Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP)

Phase III

Report on the General Health and Medical Study

September 2004

ISBN 0-9751990-3-X

Images – front cover

Title: Exercise Red Flag
Caption: As the sun rises in the desert, personnel from 1 Squadron (SQN) conduct a morning Foreign Object Damage (FOD) walk. As part of Exercise Red Flag 02, the SQN had five F-111s deployed at Nellis Air Force Base near Las Vegas, Nevada, USA.
Date: 29/8/2002
Photo: SGT Glen McCarthy 82WG Photo Section
Source: http://www.defence.gov.au/raaf/

I How to read this report

For the lay/non-technical reader

For the general background to the study and an overview of the study's structure, read Chapters 1 and 2. Chapters 3 to 6 can be skipped: they contain specifics about the recruitment of subjects, determination of exposure, measures, and analyses. For the results given in Chapters 9 to 16, the introduction and discussion sections can be read in isolation for a quick synopsis of the results in non-technical language. There is also a synopsis box at the beginning of each chapter that summarises the main points of the chapter. An overall discussion of the entire results (i.e. how different findings "cluster" together) is presented in Chapter 17.

For the technical reader

The overview of the study in Chapter 2 is supplemented by additional details regarding recruitment and characteristics of participants (Chapter 3), determination of exposures (Chapter 4), an overview of measures chosen for each domain tested (Chapter 5), and the common template for analysis and presentation of results (Chapter 6). Results for each domain are presented in an identical manner in each of the chapters from 9 to 16. In each of these chapters, the background rationale is summarised in an introduction, measures are fully described, potential confounders are listed, and any particulars about the analysis that deviate from the template are discussed. The full results are given in the Results section with additional tables in the Appendices (available on CD). Because these tables are taken directly from SAS output to avoid any transcription errors, the reference group is the DSRS group, and odds ratios are for the comparison groups relative to DSRS.

This means that harmful effects of DSRS are framed as decreased odds ratios for the comparisons. These results are described with the epidemiologist or statistician in mind. The discussion section reverses the odds ratios to a more intuitive form (i.e. harmful DSRS effects are quoted as odds ratios above 1) and summarises the results in non-technical language. Chapter 17 presents cross-tabulations between results across all domains and presents an overall discussion of all associations between positive findings for the exposed group.

II Authors

The University of Newcastle Research Associates (TUNRA) Ltd and Hunter Medical Research Institute (HMRI)

Chief investigators

- Associate Professor Catherine D'Este, DipAppSci (Medical Radiography), BMaths, DipEd, DipMedStats, PhD. Associate Professor in Biostatistics, Centre for Clinical Epidemiology & Biostatistics, University of Newcastle.
- Dr John Attia, BSc (Physiology), MSc (Clinical Epidemiology), MD, PhD (Medical Genetics), FRCP(C). Senior Lecturer in Epidemiology, Centre for Clinical Epidemiology & Biostatistics, University of Newcastle. Academic Consultant, Hunter Area Health Service.
- Dr Anthony Brown, MB BS, MPH, FAFPHM, FAFOM. Director, Population Health, Macquarie Area Health Service. Conjoint Associate Professor, Environmental and Occupational Health, University of Newcastle.
- Professor Julie Byles, BMed, PhD. Director, Centre for Research and Education in Ageing (CREA), Faculty of Health, University of Newcastle.

Associate investigators

- Associate Professor Peter W. Schofield, BSc, MBBS, MSc (Epidemiology), MD, FRACP. Clinical Director of the Neuropsychiatry Service and Senior Staff Specialist in Neurology, Hunter Area Health Service. Conjoint Associate Professor of Psychiatry, University of Newcastle.

- Associate Professor Robert Gibberd, BSc (Hons), PhD. Associate Professor, Centre for Clinical Epidemiology & Biostatistics, University of Newcastle.
- Mr Steve Lee, BSc (Hons), MSc (Clinical Neuropsychology). Senior Clinical Neuropsychologist, Neuropsychiatry Service, Hunter Area Health Service.

TUNRA Ltd

- Dr Soozy Smith, PhD. CEO, TUNRA Ltd, University of Newcastle.

Project support

- Ms Meredith Tavener, BAppSci (Hons), GradDip (Health Promotion), MmedSci. Project Manager, Centre for Clinical Epidemiology & Biostatistics, University of Newcastle.
- Mr Richard Gibson, BSc, DipEd, DipMedStats. Associate Lecturer in Biostatistics (Research), Centre for Clinical Epidemiology & Biostatistics, University of Newcastle. Project statistician.
- Mrs Maya Guest, BOH&S, BMedSci (Comm. Health)(Hons). Research Higher Degree candidate, PhD Fellow for SHOAMP.

IV Executive Summary

Background

In 1963, Australia ordered 24 General Dynamics (GD) F-111 aircraft from the United States of America. Unlike many other aircraft, the F-111 has fuel tanks that do not contain internal bladders; therefore the joints and mating surfaces in the aircraft's structure need to be sealed to prevent fuel leaks. The original sealant proved inadequate to the task, and significant fuel leaks became apparent soon after delivery of the aircraft was taken. The original sealant had to be removed (desealing) using chemical and physical methods (e.g. water jets, hand tools), before new sealant could be put in its place (resealing).

Four F-111 formal fuel tank Deseal/Reseal (DSRS) programs were implemented over two decades (1975-1999). DSRS Program 1 ran from October 1975 to December 1982. The Wing Program, used for maintaining the wing fuel tanks, was conducted from August 1985 to June 1992. DSRS Program 2 ran from February 1990 to August 1993. The Spray Seal Program ran from March 1996 to November 1999. These programs are the focus of the present study. DSRS activities were also undertaken in an *ad hoc* manner, in so-called "pick and patch" repairs, although these are not included in this study.

In early 1999, concerns were raised by the officer in charge of the aircraft maintenance section at RAAF Base Amberley about various symptoms being experienced by workers in the F-111 Spray Seal Program: the symptoms included memory loss, fatigue, and other neurological problems. As a result, the Spray Seal Program was suspended, and in January 2000 an internal investigation into the F-111 DSRS programs was conducted. The investigation concluded that a significant number of personnel had presented with symptoms consistent with solvent or isocyanate exposure and had potentially been exposed throughout all the DSRS programs. Consequently, on 19 July 2000, the Chief of Air Force appointed a Board of Inquiry (BOI) to conduct an investigation into the effects on Air Force maintenance workers of possible chemical exposure during all RAAF F-111 fuel tank repair programs, dating back to 1975.

The BOI noted deficiencies in the Occupational Health and Safety (OH&S) procedures as well as inadequate reporting of incidents and hazards, and supported the conduct of an epidemiological investigation into the health of F-111 DSRS workers. The investigation – the Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP) – aimed to assess whether adverse health outcomes reported by DSRS personnel were associated with their involvement in DSRS programs or activities.

The SHOAMP was undertaken in three phases. The first phase involved a literature review of the evidence of possible associations between chemical exposure and health outcomes, a qualitative study of a sample of those involved in DSRS, and the development of a protocol for conducting a General Health and Medical Study. The second phase involved mortality and cancer incidence studies, the last of which estimated that on the balance of probabilities there was approximately a 50% increase in cancer in the F-111 DSRS group, which was of borderline statistical significance. The third phase (the current study) is a General Health and Medical Study (hereafter referred to as the Health Study).

Methods

Workers involved in F-111 DSRS activities were identified through BOI lists, interviews, media notices, a telephone hotline, and reviews of squadron photos and records. Workers' level of potential exposure was based on a self-completed questionnaire assessing the duration and types of DSRS activities they had been involved in. Two comparison groups were then chosen as follows:

a) Technical personnel at RAAF Base Richmond (New South Wales) serving between 1975 and 1999. The purpose of this comparison group was to assess the effect of DSRS-specific exposures over and above other exposures involved in the technical musterings.

b) Other personnel, not involved in technical duties, posted at RAAF Base Amberley (Queensland) serving between 1975 and 1999. The purpose of this comparison group was to assess the effect of DSRS-specific exposures, over and above any other local exposures at Amberley, experienced by personnel not involved in aircraft maintenance.

Consenting subjects from the F-111 DSRS group and the two comparison groups were asked to complete a mailed Postal Questionnaire and undergo physical examinations and

interviews at Health Services Australia (HSA) centres. Data were collected on a number of outcomes such as:

- general health and well-being (including SF-36 quality of life)
- cardiovascular health (symptoms and postural hypotension)
- respiratory health (symptoms and spirometry testing)
- skin and breast (including dermatitis and gynaecomastia)
- neurological outcomes (including vibration sensation, colour vision, and olfaction)
- male sexual function and female reproductive health
- mental health (including depression and anxiety as measured by the Composite International Diagnostic Interview and neurasthenia)
- cognition and memory (as measured by a battery of neuropsychological tests).

Analysis

Analyses were conducted for each study outcome. Continuous variables with a non-normal distribution were transformed or dichotomised, and categorical variables with more than two categories were also dichotomised. Variability across HSA centres was examined using a test of heterogeneity (analysis of variance for continuous variables, and the Breslow-Day test for dichotomous variables). Results for each outcome are presented as two tables:

a) a descriptive table which provides summary information by group

b) a summary regression analysis table which includes (i) results for the primary analysis of the exposed group versus the two comparison groups, and (ii) results for the secondary analyses including Program 1 and Program 2 sub-groups, and the dose-response relationship.

Results

Overall, participation rates were 77%, 48% and 40% for the DSRS, Richmond and Amberley groups, yielding 659, 600 and 495 participants respectively (based on a denominator which excluded deceased individuals). However, a large proportion of the sample could not be contacted: 5% of the DSRS group, 22% of the Richmond group, and 26% of the Amberley comparison group. Excluding these individuals from the denominator (in addition to deceased individuals) gave consent rates of 81%, 62% and 54% for the DSRS, Richmond and Amberley groups respectively.

General health and well-being

On average, the F-111 DSRS group reported nearly twice the number of poor health symptoms than the comparison groups. The DSRS group recorded significantly poorer quality of life than both comparison groups on both the physical and mental component scores of the SF-36 survey.

Cardiovascular health

The Health Study focused on three potential cardiac effects of organic solvents: (a) palpitations, (b) postural drop in blood pressure, and (c) coronary heart disease. There were no differences in pulse rate, percentage of participants with hypertension, or blood pressure drop between the three groups, although baseline systolic blood pressure was slightly higher in the Amberley group (130.5) compared to the exposed group (128.0). All self-reported symptoms of dizziness, feeling faint when standing, chest pain or irregular/rapid heart beat were consistently and significantly elevated in the DSRS group compared to Amberley and Richmond, with an odds ratio (OR) of 2.0-2.5. These symptoms were statistically significant for Programs 1 and 2 and showed a dose-response effect.

Respiratory health

Based on pre- and post-Ventolin® lung function results (i.e. spirometry), only five people were classified as having reactive airways disease (i.e. asthma-like symptoms). Self-reported physician diagnoses of obstructive lung disease (i.e. bronchitis and emphysema) were significantly elevated in the DSRS group (OR=2.0), and this was congruent with the two-fold elevation in self-reported symptoms of shortness of breath and wheezing in the exposed group versus the comparisons. However, there were no differences in the spirometry results ($\mathrm{FEV}_1/\mathrm{FVC}$) at the health examination.

Dermatological and breast abnormalities

Skin conditions of interest included lipoma, dermatitis, psoriasis, and pigmented or sun-related skin lesions, together with self-reported skin irritation, dermatitis, eczema, psoriasis, and previously-diagnosed malignant melanoma. There was a strong and statistically significant two- to three-fold increase in the odds of dermatitis in the F-111 DSRS group, and this was consistent between the different methods of assessment (self-reported rash, self-reported previous physician diagnosis, and diagnosis during the health examination). There

---

® Generic name: salbutamol sulfate, Manufacturer: GlaxoWellcome.

was a less robust two-fold increase in the odds of pigmented or sun-related lesions in the DSRS group versus both comparison groups. Other outcomes were either too rare or too variable to be analysed or they showed minimal difference between groups.

Neurological outcomes

There was a two- to three-fold increase in the odds of subjective sensory and motor neuropathic symptoms in the DSRS group relative to both comparison groups, but this was not accompanied by any differences in the vibration sense tests. There was a slight increase in impaired colour vision in the DSRS group versus the Richmond group, although this was of borderline significance. There was no detectable objective change in olfaction, although there was an increase in self-reported sensitivity to smells (OR=2.5).

Male sexual function and female reproductive health

The 15-item International Index of Erectile Function (IIEF) for males identified significantly higher levels of erectile dysfunction in the DSRS group (OR=2.5). This result was consistent in subgroup analyses and showed a significant dose-response effect. There was no statistically significant evidence of any association between DSRS and miscarriage or stillbirth for female partners of male participants. There was also no detectable difference in reported difficulties getting pregnant, or in seeing a fertility specialist.

Mental health

Mood disorders (symptoms of depression and anxiety) were assessed using a variety of methods: a computerised assessment program (CIDI) administered by a psychologist, validated self-completed questionnaires (Kessler 10-item and General Health Questionnaire, GHQ 12-item), self-reported diagnoses of depression, anxiety or other somatic symptoms, and current medications (from the Postal Questionnaire). There was a fair level of agreement, at an individual level, between the self-reported indicators of mood disturbances and the objective tests administered during the health examination. The DSRS group was approximately twice as likely to report a previous diagnosis of depression and/or anxiety, to use anti-depressant medications, or to score positively on the CIDI depression and CIDI anxiety scales. Results were strong and consistent in that they were significant in the overall analysis, in both Program 1 and 2 subgroup analyses, and showed some evidence of a dose-response relationship. Data from the Kessler and GHQ also indicated that the DSRS group was at higher risk of mental distress and social dysfunction than both the comparison groups and the Australian population in general.

Neuropsychological outcomes

A comprehensive assessment of cognition was performed, including tests of executive functioning, psychomotor speed, attention/working memory, visuospatial, and new learning. The exposed group scored significantly lower on all four tests of executive functioning. All three tests of psychomotor speed indicated a statistically significant decrease in performance for the DSRS group. In the three tests for new learning/memory, differences were present between the exposed and comparison groups, with the DSRS performing worse than Richmond in the auditory verbal learning test and in the immediate and delayed recall and total learning tasks. There were no significant group differences in either of the attention/working memory tests. All these results were somewhat weakened by the lack of significance against one or other comparison group and/or subgroup analysis and the inconsistent dose-response effect. Self-reported memory complaints were significantly greater in the DSRS group relative to both comparison groups. This was consistent across Programs 1 and 2, and showed a dose-response effect.

Associations

Comparing the positive findings in the DSRS group across the various domains indicated that, in general, there was very little overlap in the neuropsychological deficits, physical and mental SF-36 component scores, or memory complaints. That is, those who scored poorly on the cognitive tests were different from those who scored poorly on the physical and mental scores on the SF-36, and were different again from those who had subjective memory complaints. There was internal consistency of the data, in that different measures of the same entity generally correlated (e.g. self-reported versus directly-observed dermatitis). There was no “typical” set of findings in the DSRS group.

Discussion

There are unavoidable uncertainties in the interpretation of the study’s results due to such factors as uncertain sampling frames, potential survivor bias, low participation rates, and multiple comparisons. Nonetheless, putting these uncertainties aside, the results point to an association between F-111 DSRS involvement and a lower quality of life and more common erectile dysfunction, depression, anxiety, and subjective memory impairment. There is also evidence, albeit less compelling, of an association between DSRS and dermatitis, obstructive lung disease (i.e. bronchitis and emphysema), and neuropsychological deficits. The exploration of causation in these findings is outside the scope and charter of this study. The *a priori* concerns regarding solvents and isocyanates cannot be fully resolved by our results.

V Acknowledgments

We would like to take this opportunity to thank members of the Scientific Advisory Committee (SAC) and Consultative Forum, and representatives from the Department of Veterans' Affairs (DVA) and the Department of Defence (DoD), for their contributions to this report.

Particular thanks goes to:

- Mrs Jean Ball and Jenny Powers, Research Centre for Gender and Health, University of Newcastle.
- Dr Geza Benke, Department of Epidemiology and Preventative Medicine, Monash University.
- Dr Lois Browne, Southeastern Area Health Service, St George Hospital, Sydney.
- Mrs Corinna Bruin, Ms Rhonda Walker, Ms Jenny Byrne, Mr Bob Kucera, Ms Debbie Quain and Ms Libby Lovett, Centre for Clinical Epidemiology & Biostatistics (CCEB), University of Newcastle.
- Dr Warren Harrex, Department of Veterans' Affairs, Defence Links Branch.
- Dr Timo Kaupinnen, Finnish Institute of Occupational Health.
- Professor Alan Mackay-Sim and Ms Leah Grant, School of Biomolecular and Biomedical Science, Griffith University, Queensland.
- Carol Pederson, School of Optometry, Queensland University of Technology.
- Dr Carol Toft and the staff of participating Health Services Australia offices.
- All members of the TUNRA data entry team.

Finally, but most importantly, a special thanks goes to the participants of SHOAMP for their time and patience. Thank you.

This page has been left blank intentionally
for formatting purposes

V List of Contents

Chapter 3 – Study Population and Recruitment

Chapter 4 – Exposure

Chapter 5 – Measures

Chapter 6 – Analysis

Chapter 7 – Quality Assurance and Ethics

Chapter 8 – Potential Confounders

Chapter 9 – General Health and Well-Being

Chapter 10 – Cardiovascular Health

Chapter 11 – Respiratory Health

Chapter 12 – Dermatological and Breast Abnormalities

Chapter 13 – Neurological Outcomes

Chapter 14 – Sexual Function and Reproductive Health

Chapter 15 – Mental Health

Chapter 16 – Neuropsychological Outcomes

Chapter 17 – Review of Findings and Discussion

VI Glossary of Terms

This section provides definitions for acronyms and terms used in this report.

***a priori* –** In the context of the General Health and Medical Study report, it means "ahead of time". That is, decisions made *a priori* were those made at the start of the study, and were not influenced by knowing the results of the study or the distribution of the data.
ABS – Australian Bureau of Statistics.
ad hoc – Latin term meaning “for this purpose”, and/or “for a specific purpose”.
ADHREC – Australian Defence Human Research Ethics Committee.
ADVISOR – A Visual Basic application known as the Analytical Database for Veteran Investigative Studies & On-line Research (or ADVISOR) system, which acted as a participant database for use by the DVA Contact and Recruitment Team to record details of eligible study subjects and make health examination appointments.
AEC – Australian Electoral Commission
AFPEMS – Air Force Personnel Executive Management System.
AIHW – Australian Institute of Health and Welfare.
AWASCo – Amalgamated Wireless Australia Serco.

Bias – Error(s) that can cause a systematic deviation from the truth.
Bio-marker – Also referred to as Biological Marker: a cellular or molecular indicator of exposure, health effects or susceptibility. Bio-markers can be used to measure internal dose, biologically effective dose, early biological response, altered structure or function, susceptibility.*
BMI – Body Mass Index.
BOI – Board of Inquiry into F-111 (Fuel Tank) Deseal/Reseal and Spray Seal Programs.

C&R Team – Contact and Recruitment Team at the Department of Veterans’ Affairs responsible for contacting potential participants of the General Health and Medical Study.
CCI – Colour Confusion Index.

---

* Source: Last, JM. A Dictionary of Epidemiology. 3rd edn. Oxford University Press; 1995, p17.

Centile – Statistics are sometimes ordered into a stepped scale of 100 evenly distributed parts. A centile is one step on the scale. For example, the 10th centile (or percentile) marks the number or value below which 10% of the data lie.
Chi-square – A test of statistical significance.
CIDI – Composite International Diagnostics Interview
Cohort – Another term for a study "group", i.e. the exposed cohort means the group of exposed individuals.
Comparison – Term used to describe those individuals whose work activities were not F-111 Deseal/Reseal related, for the purposes of the General Health and Medical Study.
Confidence interval (CI) – An interval used to estimate the likely size of a population parameter, giving an estimated range of values (calculated from a given set of sample data) that has a specified probability of containing the parameter being estimated.
Confounder – A factor or variable that can be both a risk factor for disease and associated with the exposure of interest.
Cox proportional hazards – The Cox proportional hazards model is the most commonly used regression model for survival data.

De-identify – To remove data from a document so that it cannot be linked to someone's identity.
De-plumb – The process of removing plumbing and pipes from within the F-111 fuel tank prior.
Deseal – Removal of sealant from integral tank surface
Deutan – The most common type of colour vision deficiency affecting mainly the green receptors. An individual with deutan loss will have trouble distinguishing blue-green from grey and red-purple.
DNA – Deoxyribonucleic Acid: the molecule that holds genetic information. It is the biochemical molecule that makes chromosomes and genes.
DoD – Department of Defence.
Domain – A broad aspect of health, as defined by the Health Study, that can be assessed by multiple tests, e.g. cardiovascular health, respiratory health, neurological health.
Dose-response – The relationship where a change in the amount, intensity, or duration of an exposure is associated with either an increase or decrease in risk of a specified health outcome.
DSRS – Deseal / Reseal.
DVA – Department of Veterans' Affairs.
Dyschromatopsias – Progressive loss of colour vision due to retinal diseases.

ELMA – A computerised Electronic Management of Medical Assessments (ELMA) diary system at Health Services Australia, utilised by the Department of Veterans' Affairs Contact & Recruitment Team during recruitment.
Enlisted – Being a currently-serving member of the Australia Defence Force.
Epidemiology – Scientific discipline studying the incidence, distribution and control of disease in a population.
Epstein Barr virus – A common virus that remains dormant in most people, which has been associated with chronic fatigue syndrome.
EQ – Exposure Questionnaire administered to exposed individuals and those who worked in close proximity to DSRS, to quantify their exposure.
Exposed – Term used to describe those individuals whose work activities were F-111 Deseal/Reseal related.

F-111 – Aircraft manufactured by General Dynamics, purchased by the Royal Australian Air Force from the United States Air Force. Also referred to as the "pig" or "aardvark". The aircraft holds two crew, and is an all-weather strike, attack and bomber aircraft.
Frequency matching – Matching is a technique used to adjust for the effects of potential confounding. Frequency matching is where study subjects are matched according to group characteristics.

GD – General Dynamics.
GHQ – General Health Questionnaire.

HE – Health Examination.
Health Study (the) – the General Health and Medical Study for the SHOAMP project.
Histogram – A graph of data distribution.
HMRI – Hunter Medical Research Institute.
HSA – Health Services Australia.
Hypothesis – A tentative statement which may be tested through research.

IARC – International Agency for Research in Cancer.
ICD – International Classification of Diseases.
In Toto – Latin term for "as a whole, absolutely, completely, without exception".*

---

* Source: Online Oxford English Dictionary, http://dictionary.oed.com (accessed August 5, 2004).

Logistic regression – A type of linear model used to estimate the relationship between an outcome and the explanatory variables when the outcome is discrete. For example, the outcome may be yes/no, or present/absent or mild/moderate/severe and the explanatory variables could be age, rank and posting period.

Mustering – job category held within the Defence Force.

NDI – National Death Index.
NH&MRC – National Health and Medical Research Council.
Non-Commissioned Officer – An non-commissioned officer (NCO) is anyone in the RAAF of the rank of Corporal (CPL), Sergeant (SGT) or Flight Sergeant (FSGT). Junior NCO refers to CPL, while senior NCO (SNCO) refers to SGT or FSGT.
Non-contactable – A potential participant for whom some contact information was available, however despite persistent efforts, personal contact could not be made.
Non-respondent – Individual who did not provide feedback regarding their participation in the General Health and Medical Study.

OH&S – Occupational Health and Safety

Pick and Patch – Repair work similar to DSRS conducted on the F-111 fuel tanks prior to, during, and after the formal Deseal/Reseal programs, involving entry into the F-111 fuel tanks, carefully locating suspect areas of sealant, and removing the sealant from the area of concern.
Pilot – In the context of the Study of Health Outcomes in Aircraft Maintenance Personnel, "pilot" refers to the Round One trial period during which project documentation and health examination procedures were tested.
PMKeyS – Personnel Management Key Solutions.
Point estimates – Results of estimation expressed as a single value.
Posting – Location of work duties assigned to an individual.
PPE – Personal protective equipment.
PQ – Postal Questionnaire.
Program 1 – One of four formal F-111 fuel tank Deseal/Reseal (DSRS) programs which were implemented over two decades (1975-1999). Program 1 at Amberley RAAF Base ran from October 1977 to December 1982. Earlier DSRS work was conducted at Sacramento from 1975.

Program 2 – One of four formal F-111 fuel tank Deseal/Reseal (DSRS) programs which were implemented over two decades (1975-1999). Program 2 ran from February 1990 to August 1993, including work conducted at Sacramento.

Protan – A vision deficiency affecting mainly the red receptors. An individual with protan loss will have trouble distinguishing red-green and will confuse red-orange with blue-green and grey.

PQ – Postal Questionnaire.

QML – Queensland Medical Laboratory.

R4 and R5 – The "R" stands for "routine" servicing, and the number designates the type of service to be carried out on the F-111 aircraft.

RAAF – Royal Australian Air Force.

Rank – Position of seniority/authority and responsibility held within the Australian Defence Force, particularly the Air Force, for the purposes of this study.

Reclassification – The process of assessing DSRS involvement by participants and providing a classification of "exposed" or "unexposed" for the purposes of the General Health and Medical Study.

Recruitment – The process of contacting each potential SHOAMP participant, providing information about the General Health and Medical Study

Regression analyses – Regression analysis is a statistical method used to examine the degree of association between two or more variables.

Re-plumb – Following resealing, the F-111 fuel tanks had the plumbing replaced and re-activated.

Reseal – Where the interior surface of the F-111 fuel tank was cleaned and fresh sealant was laid.

Respondent – Individual who provided feedback regarding their participation in the General Health and Medical Study.

Risk ratio – A type of measure of relative risk. It is the ratio of the incidence rate in the exposed group to the incidence rate in the unexposed group.

Round one – The pilot recruitment phase, during which all contact and health examination processes were trialed.

Round two – The main body of data collection for the General Health and Medical Study.

Round three – Continuing data collection for the Health Study including newly reclassified individuals.

SAC – Scientific Advisory Committee.
SAS – Statistical software.
SF-36 – Short Form 36-item quality of life survey.
SHOAMP – Study of Health Outcomes in Aircraft Maintenance Personnel.
Solvent – A substance, usually a liquid, in which other substances are dissolved.
Spray Seal – One of four formal F-111 fuel tank Deseal/Reseal (DSRS) programs which were implemented over two decades (1975-1999). The Spray Seal Program ran from March 1996 to November 1999.
SR51 – A chemical desealant used in F-111 DSRS Program 1.
STATA – Statistical software.

Tertile – Division of data for a variable into thirds.
Tritan – An individual with a tritan deficit will confuse violet with grey and yellow-green. Tritan loss is rarely inherited, and shows a 'confusion axis' from yellow to blue.
TUNRA – The University of Newcastle Research Associates.

USA – United States of America.

Variable – A variable is a characteristic that is being investigated in a research study that differs from subject to subject and/or from time to time.

Weighting – The process by which data are adjusted to reflect a known population or comparison profile. A "weight" is the value assigned to a particular criterion.
WHO – World Health Organisation.
Wing DSRS – One of four formal F-111 fuel tank Deseal/Reseal (DSRS) programs which were implemented over two decades (1975-1999). The Wings DSRS Program ran from August 1985 to June 1992.

Z-Score – A measure of the distance in standard deviations of a sample from the mean. Calculated as (X - the mean of X) / standard deviation of X.

1 Background

Chapter summary

This chapter provides an introduction to the F-111 aircraft Deseal/Reseal (DSRS) fuel tank maintenance programs, which were implemented over more than two decades (1975-1999). Each program involved different work processes and chemical substances, with concerns being raised about various symptoms experienced by DSRS workers as a possible direct result of these exposures. In 2000, a Board of Inquiry (BOI) was appointed to investigate the chemical exposure of the DSRS workers, with support for the conduct of an epidemiological study of the health outcomes of F-111 DSRS workers. Numerous outcomes were of concern, including mortality and cancer incidence (reported previously), neurological and neuropsychological outcomes such as loss of memory and cognition, anxiety and depression, and loss of colour vision.

Chapter contents

1.1 Introduction

In 1963, Australia ordered 24 General Dynamics F-111 aircraft from the United States of America. The aircraft, not yet constructed, were due to be delivered by 1968; however, a series of problems deferred production, so the final (24th) aircraft was not delivered until October 1973.

The Royal Australian Air Force (RAAF) describes the F-111 as a long-range, strategic strike and reconnaissance weapons platform, which forms a significant component of Australia's defence capability. To maximise operational range and endurance, the F-111 makes the best use of all available space for fuel: it carries fuel in the wings and the fuselage and in external fuel tanks.

Unlike many other aircraft, the F-111 has fuel tanks that do not contain internal bladders. This means that there is no barrier between the fuel and the aircraft's internal metal surfaces. Consequently, the joints and mating surfaces in the aircraft's structure need to be sealed to prevent fuel leaks. These sealing systems use complex chemical sealant formulations and applications. The sealants are required to cope with extreme environments including heat generated during supersonic flight, structural strain as a result of manoeuvring, and the chemically-hostile environment of being immersed in aviation turbine fuel. The original sealant proved inadequate to the task, and significant fuel leaks became apparent soon after the aircraft had been delivered.

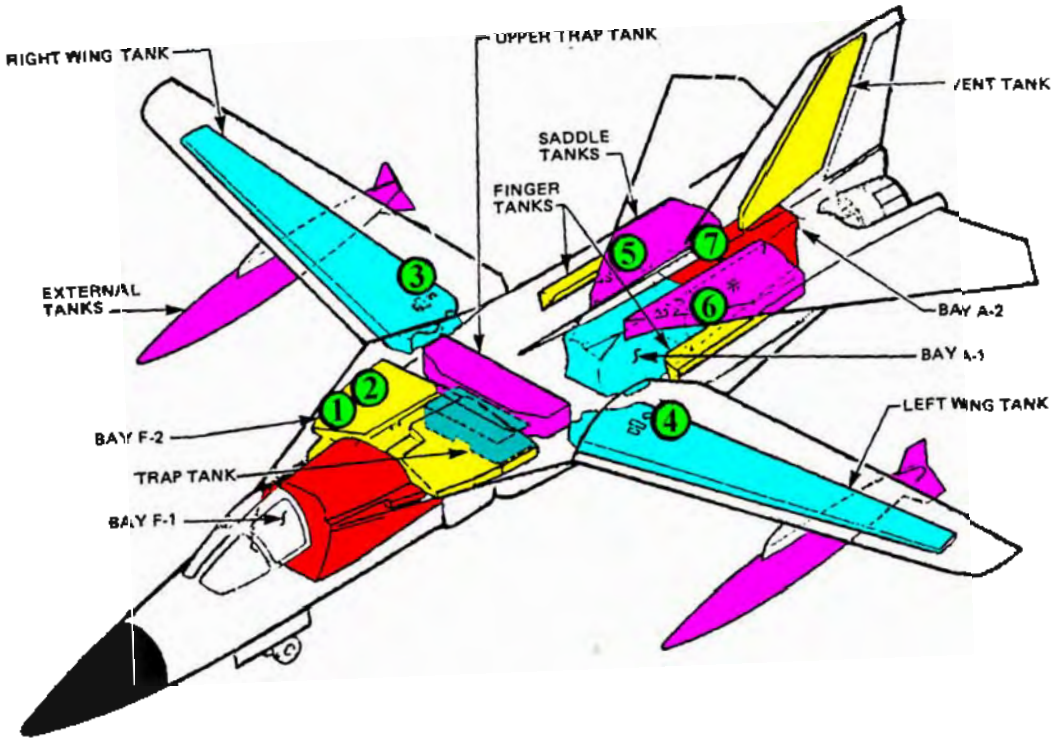
1.2 F-111 fuel tank repairs and maintenance

F-111s have numerous fuel tanks of different sizes (Figure 1.1).* At the time of manufacture, sealant was put between the overlapping metal surfaces in the tanks. Over time, the sealant began to deteriorate and a variety of repairs and maintenance procedures were implemented, both on and inside the F-111 fuel tanks, to correct these problems. In particular, the repair of sealed joints was necessary to correct the fuel leaks and it was also a part of routine maintenance programs. This involved firstly removing the original sealant

* Figures 1.1 to 1.4 printed with the permission of the Department of Defence, from the Report of the Board of Inquiry into F-111 (Fuel Tank) Deseal/Reseal and Spray Seal Programs.

inside the fuel tanks (desealing) and then replacing it with new sealant (resealing). Removal of the original sealant required firstly the use of chemicals, secondly water jets, and finally hand tools (manually). For the fuselage fuel tanks the process of desealing and resealing required physical entry to the tanks, while for wing tanks this was not required. Four formal F-111 fuel tank Deseal/Reseal (DSRS) programs were implemented over more than two decades (1975-1999), each involving different processes. Following training and preparatory works (including work conducted at Sacramento, USA) from 1975 to 1977, Program 1 ran from October 1977 to December 1982. The Wing Program, used for maintaining the wing fuel tanks, was conducted from August 1985 to June 1992. Program 2 ran from February 1990 to August 1993, again including work conducted at Sacramento. The Spray Seal Program ran from March 1996 to November 1999. The processes, activities and occupational titles involved in each of these four DSRS programs are outlined in Appendices 1A and 1B.

Figure 1.1 : F-111 with all fuel tanks highlighted



Three DSRS programs - Program 1, Program 2 and the Spray Seal Program – involved physical entry into the F-111 fuel tanks. For the Wing Program the tanks were open to the air, so there was no need for personnel to be completely inside the wing tank (Figure 1.2). Specific preparatory tasks had to be undertaken prior to entering the tank. The process involved removing all fuel from the tanks (defuelling and depuddling) and then removing plumbing and pipes (deplumbing). Once inside the fuel tank, maintenance personnel manually removed the defective sealant from the surface of the tank (desealing) (Figure 1.3 and Figure 1.4). The surface was then cleaned and fresh sealant was laid (resealing). The tanks were then replumbed (replacing plumbing) and refuelled. The methods used for these processes changed over time. Spray sealing, a different process for the main fuselage tanks, did not require any sealant to be removed unless it was obviously defective. Instead, the surface of the old sealant was cleaned and prepared, and a new coat of sealant was sprayed directly over the old sealant.

Figure 1.2 : Aircraft technicians working on the Wing Program

Figure 1.3 : A technician checking sealant in the fuselage tank

Figure 1.4 : An aircraft technician using a dental mirror and pick to remove sealant

The Wing Program was developed in response to fuel leaks within the aircraft wings, which, until 1985, were repaired in an *ad hoc* manner. As wing tanks were too small for maintenance personnel to enter, the top wing plank was removed from the aircraft and the tank completely opened so that work could be conducted from outside. Desealing was conducted using water picks and walnut-shell blasting. The Wing Program also utilised contracted civilian workers from Amalgamated Wireless Australia Serco (AWASCo), under RAAF supervision.

Through the early- to mid-1980s a series of maintenance programs was conducted on fuel tanks other than those in the fuselage and wing areas (i.e. the vent tanks and weapons bay tanks). This activity was conducted between the formal DSRS programs and involved a working environment similar to the Wings Program. In addition, aircraft undergoing routine servicing types (R4 and R5 for the F-111; see Glossary) in between the formal DSRS programs had individual fuselage fuel tank repairs, as did some aircraft allotted from the operating squadrons.

In the 1990s the United States Air Force (USAF) began using the new spray seal process to repair F-111 fuselage tanks. This process significantly reduced the time taken to repair the leaks: it consisted of ‘water pick’ desealing and cleaning, followed by an overcoat application of the sealant. The process was further refined to do away with the pick process and to use only patch repairs, with the new sealant being sprayed over the old one. This process was used by the RAAF from 1996 until early 2000, after which it was suspended.

Workers on the DSRS programs spent a significant proportion of their time inside the fuel tanks in conditions that were cramped, almost fully enclosed and inadequately ventilated. They frequently worked in this environment for extended periods of time, sometimes up to five hours. The risk of exposure to jet fuel and to the chemicals used in the DSRS processes required the use of personal protective equipment (PPE). The protective suits worn during the Spray Seal Program were unpleasant and additionally hazardous due to their weight and bulk and because of the impervious nature of their material; also, they increased physical demands and interfered with thermo-regulation. Following an increase in the number of health complaints by Spray Seal personnel, it was found that the protective overalls did not provide adequate protection against the chemicals, so that workers wearing them came into direct contact with the chemicals in liquid and vapour form.

1.2.1 Other activities associated with F-111 DSRS

In addition to those tasks recognised as being directly associated with the F-111 DSRS programs (described above), there were also several other related activities conducted by RAAF personnel which could have involved exposure to chemicals during DSRS procedures. These associated activities included the mixing of DSRS chemicals and the disposal of DSRS chemicals, as well as tasks carried out by personnel of other mustering types that involved periods of work in close proximity to the DSRS processes and *ad hoc* repairs conducted outside formal programs.

1.2.1.1 Disposal, storage and mixing

Health and safety issues have been raised about the methods used to dispose of some of the chemicals in the DSRS programs – in particular, the process for disposing of the chemical SR51 (a chemical desealant used in Program 1). Disposal techniques included incineration and storage of waste in drums. The Board of Inquiry (BOI) found that both these techniques were not properly conducted or were inappropriate as a means of disposal. The incinerator was found to have been malfunctioning for an unknown period, possibly years, and the storage drums were found to be leaking on more than one occasion. There was also evidence that SR51 was regularly (i.e. a couple of times a week) burnt during fire-crew training from the late 1970s until approximately 1990.

Mixing of the sealant was another activity associated with potential exposure to toxic substances. However, in simulated workplace environments, testing showed that the level of exposure to airborne contaminants during mixing (while conducted in an open work area) was below the Australian Workplace Exposure Standard.

1.2.1.2 Pick and patch

Some repair work similar to DSRS was conducted on the F-111 fuel tanks prior to, during, and after the formal Deseal/Reseal programs. The operation known as “Pick and Patch” was used to repair F-111 fuel tanks that were leaking. As with the formal DSRS programs, the Pick and Patch process involved entering the F-111 fuel tanks, carefully locating suspect areas of sealant, and removing the sealant from the area of concern plus a margin around it using solvents and tools such as dental picks. A patch of new sealant would then be applied. The aircraft subject to this process were in operational squadrons. As such, the Pick and

Patch process involved running (*ad hoc*) repairs by the best means available whenever needed – and with a sense of urgency given the requirements for a certain number of aircraft to meet flying commitments at any one time. It also appears that many aircraft other than the F-111 were subject to this procedure, and in some locations a number of maintenance staff were involved for various lengths of time. It is recognised that some individuals may have spent more time working on Pick and Patch than on the formal DSRS programs.

1.2.1.3 Other contact

Other groups which have been defined as possibly exposed to DSRS substances include disposal crews, other flight maintenance crews, fire crews, and RAAF members borrowed from other units during staffing shortages.

1.2.2 Health concerns

Despite the use of various forms of respiratory equipment and protective clothing, it is probable that DSRS workers were exposed to a variety of chemicals. The DSRS methods used during each program varied, and Occupational Health and Safety (OH&S) requirements for respiratory and skin protection may not always have been complied with. It is also not clear if respiratory equipment was used in the first program.

In early 1999 concerns were raised by the officer in charge of the aircraft maintenance section at RAAF Base Amberley about various symptoms experienced by workers in the F-111 Spray Seal Program. The symptoms included memory loss, fatigue, and other neurological problems. Staff had been concerned over the possible connection between these symptoms and the F-111 DSRS work since late 1998. However, a visit by two affected workers to the Amberley Base medical centre ‘failed to produce a response’. The officer in charge requested that tests on the personnel in question be carried out ‘as deemed appropriate by medical staff’. A full blood count, liver function tests and urinalysis were conducted on affected personnel. The results were inconclusive and the medical centre discounted any association between the symptoms and the DSRS programs. In September 1999 a sergeant, new to the fuel tank repair section, became concerned about the same symptoms and encouraged the affected workers to report to the medical centre. The medical staff then referred the matter to a higher level at the Base. As a result, the Spray Seal Program was suspended in January 2000.

1.2.3 Response to concerns

1.2.3.1 Internal investigation

In 2000, an internal investigation into the F-111 DSRS programs concluded that a significant number of personnel had presented with symptoms consistent with solvent or isocyanate exposure. This conclusion was based on the very strong presumptive evidence that workers had been exposed and their symptoms were consistent with that exposure. It was found that exposure could have arisen from:

- entering tanks without adequate protective equipment
- exposure while removing protective equipment
- use of inappropriate or inadequate protective equipment for certain procedures
- failure to comply with procedures
- inadequate ventilation
- exposure to very high concentrations of chemicals, by kneeling in puddles for example.[1]

The investigation found that this was not the first time concerns had been raised over adverse health effects associated with the DSRS programs. There were at least four documented incidents where workers in the first program had exhibited symptoms consistent with chemical poisoning. None of these incidents had resulted in investigations being made into the work procedures associated with the programs. It was also apparent that there was potential for exposure to have occurred during all of the DSRS programs. In the face of the overwhelming potential health and legal consequences, a Board of Inquiry (BOI) was appointed to investigate the chemical exposure of the F-111 DSRS workers and to do so, where possible, in full public view.

1.2.3.2 Board of Inquiry

On 19 July 2000, the (then) Chief of Air Force appointed a BOI to conduct an inquiry into the effects on Air Force aircraft maintenance workers of possible chemical exposure during all RAAF F-111 fuel tank repair programs, dating back to 1975.

The terms of reference for the BOI described its main role as making inquiries into, and findings and recommendations concerning, the following matters:

a) the Deseal/Reseal Program conducted by the RAAF and/or contractors* in the late 1970s and early 1980s on F-111 aircraft at 501 Wing Amberley (Deseal/Reseal Program 1)

b) the Deseal/Reseal Program conducted by the RAAF and/or contractors in the early 1990s on F-111 aircraft at 501 Wing Amberley (Deseal/Reseal Program 2)

c) the Wing Tank Deseal/Reseal Program conducted by the RAAF and/or contractors in the late 1980s and early 1990s on F-111 aircraft at 501 Wing Amberley (the Wing Program)

d) the post-1996 Spray Seal Program conducted by the RAAF and/or contractors on F-111 aircraft at 501 Wing Amberley (the Spray Seal Program).

Among other factors, the BOI found that there had been deficiencies in the OH&S procedures as well as inadequate incident- and hazard-reporting,. These deficiencies had potentially resulted in workers involved in the DSRS programs from 1975 to 1999 being exposed to toxic chemicals and suffering adversely as a consequence. The BOI's report supported the conduct of an epidemiological study of the health of F-111 DSRS workers.

1.2.3.3 Epidemiological study

In order to determine if there was evidence to support any chronic, long-term adverse health problems of personnel involved in the DSRS programs, an epidemiological investigation was commissioned by the Department of Defence (and administered by the Department of Veterans' Affairs). The aim of this investigation – the Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP) – was to assess whether adverse health outcomes were associated with Deseal/Reseal programs or activities. Numerous outcomes were of concern, including mortality and cancer incidence, and neurological and neuropsychological outcomes such as memory loss, cognitive impairment, anxiety and depression.

* The BOI stated that both RAAF members and contractor staff had been employed at various stages across the four DSRS programs, with Hawker De Havilland conducting the second DSRS program under contract and AWASCo providing contract labour staff to supplement Air Force personnel working on the Wing DSRS Program (BOI Vol 2, Part 1, Chapter 12).

1.2.4 F-111 Deseal/Reseal Interim Health Care Scheme

The F-111 Deseal/Reseal Interim Health Care Scheme was instituted on 8 September 2001 to ensure that appropriate health care would be available to RAAF and civilian personnel who may have suffered adverse health effects following exposure to F-111 DSRS-related activities. The Scheme administers payments related to two categories of participant:

1) 'Group 1' is defined as those persons engaged in the F-111 DSRS programs (including persons exposed to chemicals as a result of those programs). To members of Group 1, the Scheme provides payment for treatment and/or counselling (including genetic counselling).

2) 'Group 2' is defined as other personnel who worked on RAAF Base Amberley at the time of the F-111 DSRS programs. Also included in this class are immediate family members of Group 1 individuals. To members of Group 2, the Scheme provides payment for counselling (including genetic counselling).

For Group 1 participants specifically, the Scheme operates as an interim system of health care, administering payments to participants during the period that their relevant compensation claims have been lodged but not determined. Subject to administrative requirements, the Scheme provides payment for treatment of Group 1 participants for a range of conditions that might reasonably be related to their work with F-111 DSRS; conditions that fall outside this range are considered on a case-by-case basis. The broad range of health conditions covered at the time of printing this report included:

- skin rashes and associated systemic conditions
- neurological conditions
- mental disorders
- personality changes
- chronic infections
- neoplasms
- haematological conditions
- liver diseases
- chronic respiratory conditions
- gastrointestinal problems
- fatigue
- coronary heart disease, its precursors and sequelae.

Group 2 participants, like those in Group 1, are eligible for genetic counselling. In most instances the number of counselling sessions for Group 2 participants that the Scheme pays for is limited to five, although additional counselling may be approved on a case-by-case basis.

1.3 Scientific Advisory Committee

In late 2000, a Scientific Advisory Committee (SAC) was appointed by the Secretary of the Department of Veterans' Affairs, Dr Neil Johnston, acting on behalf of the (then) Minister for Veterans' Affairs, the Hon. Bruce Scott. The role of the Committee was to oversee scientific aspects of the proposed epidemiological Study of Health Outcomes in Aircraft Maintenance Personnel and to act as arbiters on any issue of science that needed to be resolved. The Committee was also to have a role in assessing the merit of the various tenderers that wished to undertake the study. The Committee is described in more detail in Chapter 2.

1.4 Request for Tender

The Commonwealth Departments of Defence and Veterans' Affairs implemented a tender process to find a suitably-qualified research organisation to undertake an epidemiological study. The terms of reference are included in Appendix 1C. A successful tender was submitted to the Department of Veterans' Affairs by The University of Newcastle Research Associates Ltd (TUNRA) on behalf of researchers from The University of Newcastle and the Hunter Medical Research Institute (HMRI), Newcastle. TUNRA provides a broad-based research and specialist consulting service to industry and the wider community. The HMRI is a virtual institute that allows autonomous research groups at different sites to be united by shared management structure, resources and philosophy. The tender by TUNRA and HMRI proposed an epidemiological study – the Study of Health Outcomes in Aircraft Maintenance Personnel – which consisted of three phases.

1) Phase I involved a detailed literature review[*] to obtain the most recent information on the relationship between exposures potentially encountered during DSRS activities and

[*] Study of Health Outcomes in Aircraft Maintenance Personnel, Phase I Literature Review, Final Report July 2003.

possible outcomes, a qualitative study* to obtain in-depth information on activities and exposures of individuals involved in DSRS, and finalisation of the definitions of DSRS exposure and comparison groups.

2) Phase II involved an interim and second mortality and cancer incidence study† based on record linkage with data from the National Death Index, National Cancer Statistics Clearing House and all State and Territory Cancer Registries.

3) Phase III is the General Health and Medical Study (hereafter referred to as the Health Study) which involves an Exposure Questionnaire, a Postal Questionnaire and a series of health and neuropsychological examinations to assess exposure and outcomes for individuals involved in DSRS activities and to make appropriate comparisons.

1.5 General Health and Medical Study

The General Health and Medical Study involved Air Force personnel who had been involved in any of the F-111 DSRS programs and related activities, as well as two comparison groups: (a) personnel from non-technical workgroups (no aircraft maintenance) at Amberley RAAF Base; and (b) personnel from Richmond RAAF Base who did carry out technical work. A number of health outcomes were studied; the overview of the study is in the next chapter.

1.6 References

1. RAAF. Chemical Exposure of Air Force Maintenance Workers, Report of the Board of Inquiry into F-111 (Fuel Tank) Deseal/Reseal and Spray Programs, volume 2. Appendix A. Doctor Donaldson's Report "Nature and Extent of Health Complaints". (Royal Australian Air Force, Canberra, Australia, 2001).

* Study of Health Outcomes in Aircraft Maintenance Personnel, Phase I Qualitative Interviews, Final Report July 2003.

† Study of Health Outcomes in Aircraft Maintenance Personnel, Phase II Mortality and Cancer Incidence Study, Interim Report, July 2003; and Study of Health Outcomes in Aircraft Maintenance Personnel, Phase II Mortality and Cancer Incidence Study, Second Report, April 2004.

2 Overview of the General Health & Medical Study

Chapter summary

This chapter provides an overview in non-technical terms of the General Health and Medical Study of the SHOAMP project (Study of Health Outcomes in Aircraft Maintenance Personnel). Details are provided of the Deseal/Reseal (DSRS) and comparison group categories and of how the study was conducted, and there is a brief overview of the ways by which health information was collected from participants (i.e. postal questionnaire and health examination). Domains (see Glossary) covered in the study included neurological function, mental health, memory and other neuropsychological tests, sexual function, and cardiovascular and respiratory health. Full descriptions are also given of The University of Newcastle Research Associates (TUNRA) study team, the Scientific Advisory Committee (SAC), the Consultative Forum, the Department of Veterans' Affairs (DVA) contact and recruitment team, Health Services Australia (HAS), and other study associates.

Chapter contents

2.1 Study aims

The aim of the General Health and Medical Study was to compare a series of general health, medical and neuropsychological outcomes between F-111 Deseal/Reseal (DSRS) personnel and appropriate comparison groups.

2.2 Research questions and hypotheses

The General Health and Medical Study aimed to answer the following research questions:

- *Is there an association between adverse health status and an involvement in F-111 Deseal/Reseal activities?*
- *If so, what is the nature and strength of those associations?*

The hypotheses of the study were that Australian Defence Force and contracted civilian personnel involved in any of the DSRS programs would have, relative to an appropriate comparison group:

- a higher prevalence of specific neurological disorders
- a higher prevalence of neuropsychological impairment
- a higher rate of adverse reproductive outcomes
- poorer general health and quality of life.

2.3 Study design

The General Health and Medical Study was designed as a retrospective cohort (see Glossary) to assess whether there are differences in adverse health outcomes between F-111 DSRS exposed and comparison personnel (see “Definition of groups” below). The study involved a mailed Postal Questionnaire in addition to a series of general health, medical and neuropsychological assessments of consenting participants from the cohort involved in DSRS activities (the “exposed” cohort) and two appropriate comparison groups (the “Richmond comparison” and “Amberley comparison” cohorts). The health assessments were conducted for the Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP)

by health professionals at Health Services Australia (HSA) centres throughout selected regions of Australia.

Outcomes for the General Health and Medical Study were defined *a priori* (see Glossary). Wherever possible validated instruments were selected for use (see Chapter 5: Measures), based upon evidence from the scientific literature. Where there was a choice of more than one relevant instrument, the most appropriate measure was selected based on length of time of administration, validity, reliability, cost, and appropriate comparison or normative data.

2.3.1 Study population and recruitment

2.3.1.1 Definition of groups

The study comprised three groups: the F-111 DSRS group (referred to from this point on as the “exposed” group) and two comparison groups. The exposed group were workers who had been involved in any of the F-111 DSRS programs and who were identified and registered on the Department of Veterans’ Affairs (DVA) F-111 list.

It was important that the comparison cohort be derived from the same source population as the exposed cohort to ensure that the groups being compared were as similar as possible with respect to all other factors except the exposure of interest. Comparison with a cohort from the general population was not considered appropriate for SHOAMP, due to the health and fitness requirements of those applying for and being accepted into the Defence Forces, in addition to lifestyle and cultural issues specific to Defence Force employment.

Due to the uniqueness of the exposure, it was appropriate to have two comparison groups, each of which was similar to the exposed group in at least some respects. Observing similar results in both comparison groups would suggest that the study results would be more likely to be valid. The first comparison group comprised Air Force personnel who worked at the same base (RAAF Base Amberley) at the time the programs were conducted, but who were involved in non-technical “musterings” (job categories in the Defence Forces). This would enable a comparison of individuals who had similar “Base” exposures but were not exposed to aircraft maintenance duties in general and to F-111 DSRS programs specifically. This group should therefore not have been exposed to chemicals or hazards inherent in any form of aircraft maintenance.

The second comparison cohort consisted of Air Force personnel posted to a different base (RAAF Base Richmond) at the time of F-111 DSRS programs or activities and who were involved in technical trades but had not been involved in F-111 DSRS activities. This would allow comparison of outcomes for F-111 DSRS individuals over and above any non-specific adverse effects of general aircraft maintenance. A chart illustrating the rationale for the comparison groups is provided in Appendix 2A. The mustering categories defining the Amberley and Richmond comparison groups are listed in Appendices 2B and 2C.

2.3.1.2 Identification of potential participants

Exposed cohort

Exposed individuals were selected from a database established by the Department of Veterans' Affairs – referred to as the F-111 list – and were initially identified by the Board of Inquiry (BOI) through Defence Force records. The identity of these individuals was determined from fuel tank repair records, RAAF posting and attachment records, and contractor staff records. Squadron photos were also used to identify people who were working on the programs. These people were then able to name co-workers who had not been identified by the previous means. During the course of the inquiry, approximately 700 people were identified as having been involved in DSRS activities to some degree. In addition the BOI identified two other groups who may have been at risk from DSRS activities. The first group was personnel employed on duties closely related to DSRS activities. These included tradespeople who carried out fuel tank repairs outside the formal DSRS programs. Although the amount of time spent repairing tanks was reduced, the amount and type of chemicals used were similar in many respects to those used in DSRS Program 2 (see Glossary) . The second group included personnel working in such proximity to F-111 DSRS activities as to be at risk. Those most evidently at risk were the Boiler Attendants whose job it was to dispose of the SR51[*] by incineration. Also part of this group were Surface Finishers who repaired the fuel tank paint as required, Electrical Fitters/Avionics Technicians who removed and then reinstalled electrical components within the fuel tanks, and Non-Destructive Inspection Technicians who performed structural inspections before tanks were resealed. Immediately before applying both primer and sealant to the fuel tanks, a number of products first had to be mixed. This task was performed using a mechanical mixer.

[*] Used in F1, F2, A1 and A2 fuselage tanks, as a chemical softening agent during DSRS Program 1. It consists of petroleum solvent (high flash aromatic) 60-90%, thiophenol 5-10%, dimethylacetamide 5-10% and triethylphosphate 1-5%.

Individuals who undertook this task were also considered to be potentially at risk of chemical exposure.

The initial list of personnel supplied by the BOI was not exhaustive, and DVA worked to expand it. A wide advertising scheme was set up to inform people who had worked on the programs about the possible health risks. A hotline was established, and advertising appeared in national daily newspapers as well as in internal Defence publications and circulars and on official web sites (see Appendix 2D). Workers who had been in contact with the DVA were asked to name anyone else they could remember who might have been involved with DSRS activities. When someone contacted the DVA hotline, their details were recorded and added to the database. DVA sought to capture the names of as many people as possible who may have been involved with the DSRS programs or worked within close proximity even though they may not have been eligible to participate in the study.

Information on the DVA F-111 list included a variety of data relating to individuals' self-reported involvement in DSRS activities – such as time of exposure, and rank at time of exposure – in addition to information on age, date of birth, and problems experienced as a result of involvement with DSRS activities.

Comparison cohorts

Data for the comparison group sampling frames were provided by DVA from the computerised Air Force Personnel Executive Management System (AFPEMS) records. (A brief description of the procedure for obtaining the comparison cohorts is provided in Appendices 2E and 2F.) Two different files were provided. They included all individuals who had been posted to RAAF Base Amberley (for comparison group one) or RAAF Base Richmond (for comparison group two) between 1 January 1975 and 31 December 1999. The comparison cohorts needed to have a similar distribution to the exposed cohort for variables thought to be associated with the outcomes:

- gender
- age group (≤19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+; based on age at 12 September 2002, the date of lockdown of the DVA F-111 list)
- rank (Enlisted, Non-Commissioned Officer, Officer)
- period of posting or DSRS involvement at the time of the fuel tank maintenance programs.

This selection process was complicated because of multiple postings for individuals. For each potentially-exposed individual, the year of first exposure was obtained from the DVA F-111 list. This was then categorised into one of five time periods: 1975-1979, 1980-1984, 1985-1989, 1990-1994 and 1995-1999.

The comparison groups were frequency-matched to the exposed group for gender, age, period of posting or exposure, and rank, at the time of the fuel tank maintenance programs. Further details regarding study population and recruitment are provided in Chapter 3.

2.3.2 Exposure

A valid measurement of exposure for SHOAMP was problematic for two reasons:

1) Involvement in F-111 DSRS programs had occurred up to 30 years ago.
2) There were no records of tasks performed and exposure details for each DSRS worker.

The process was further complicated by the fact that the F-111 DSRS programs and related activities involved a substantial number of potentially detrimental exposures, including an estimated 60 hazardous substances (see Appendix 2G) as well as the environment in which the work was conducted (e.g. confined spaces, high temperatures). In addition to the independent effects of such exposures, combinations of exposures may have had a multiplicative or synergistic effect. It was recognised very early in the design of SHOAMP that it would not be possible to assess the effect of individual exposures or hazards, nor to attribute any adverse health outcomes to specific components of DSRS (i.e. health problem X was caused by chemical Y). At best, the effect of individual programs, as a subset of all DSRS activities, could be investigated, with the understanding that this assessment would be confounded by time. Thus the “exposure” to be assessed for SHOAMP was actual involvement in F-111 DSRS rather than a measure of specific components of DSRS. This involvement was initially divided by DVA personnel into the following three “loose” categories.

- *Exposure Category 1.* Workers fell into this category if they were at any time directly involved with the Amberley F-111 Deseal/Reseal programs.
- *Exposure Category 2.* This category included workers who were not directly involved with the DSRS programs but worked in such proximity as to be at risk.
- *Exposure Category 3.* These individuals were at RAAF Base Amberley during the exposure period of interest (i.e. conduct of the DSRS programs) but were not directly involved and did not work within such proximity as to be at risk.

This assignment of categories was based on a semi-structured interview (see Appendix 2H). Workers were additionally categorised according to the DSRS program or activity on which they worked: for example Program 1, Program 2, Wing Program, Spray Seal Program, chemical disposal, mixing and/or storage. This assessment included searching service records, making follow-up contact with individuals, and checking other available records. The classification of exposure by DVA was not considered to be consistent or detailed enough to be definitive, and a more precise assessment of exposure was required. Thus, as part of the mail-out to potential participants, an Exposure Questionnaire (EQ) was included (see Appendix 2I).

Due to the remoteness of some of the programs (approximately 30 years ago) and the incompleteness of the records, it was difficult to ascertain exactly how many people had worked on the programs, and it was unclear exactly how complete the DVA F-111 list was. The EQ was used to systematically classify each participant according to:

- whether they had been involved in any of the four F-111 DSRS programs (Program 1, Program 2, Wing Program and/or Spray Seal Program) or associated activities (chemical storage, mixing and/or disposal)
- which activities they had performed
- the length of time they had spent in a program (this determines the “dose” of exposure).

The method for estimating this level or “dose” of exposure is discussed more fully in Chapter 4 (Exposure).

2.3.3 Health outcomes measures

This section provides an overview of the measures used for the General Health and Medical Study. A more detailed description of the health outcome measures is provided in Chapter 5. As discussed previously, some outcomes were defined *a priori* whereas others were chosen after the literature review was conducted (see Table 2.1). Some of the outcomes were considered of primary interest and are reported here. Other outcomes were of secondary interest and will be reported later. Other outcomes were judged to be clinically important as part of a general physical examination and were included for the benefit of the participants.

The results of those outcomes that were judged to be important as part of a general physical examination were communicated directly to participants and their nominated medical

practitioner (see Appendix 2J for an example of the summary feedback letter sent to participants following their health examination). Two different Postal Questionnaires were given to potential participants: one for males and one for females (see Appendix 2K for the male version of the Postal Questionnaire and Appendix 2L for the sections that differed in the female version). A separate, sealed Female Partner Questionnaire for female partners of male study invitees was also included (see Appendix 2M). One copy of the questionnaire was provided to each potential male participant as part of their initial mail-out, with the offer of additional questionnaires should they be required. Appendix 2N details the Health Examination Booklet, where HSA personnel recorded all health assessment results for each participant.

Table 2.1 : Outcome measures

PRIMARY OUTCOMES		
Domain	**Postal Questionnaire**	**Health Examination**
General health and wellbeing	Self-reported symptoms experienced from a list of 80 items. Admission to hospital. Self-completion of the Medical Outcomes Survey Short Form 36-item quality of life survey.	Urinalysis, screening blood work including renal function tests, liver function tests, electrolytes and full blood count.
Cardiovascular health	Self-reported chest pain, rapid/pounding/irregular heartbeat, dizziness, fainting, blackouts and/or feeling like fainting when standing up. Self-reported previous physician diagnosis of high blood pressure and/or heart disease.	Blood pressure and postural drop in blood pressure, measured by HSA doctor.
Respiratory health	Self-reported shortness of breath and/or wheezing. Self-reported physician diagnosis of asthma, bronchitis, pneumonia, tuberculosis and/or emphysema.	Spirometry, pre-Ventolin, and post-Ventolin as conducted by nurse.
Dermatological and breast abnormalities	Self-reported symptoms of rash or skin irritation, dermatitis, eczema and/or psoriasis. Self-reported experiences of “any new lump(s) in the breast area”, “any change to the skin of the nipple/breast”, “an unusual increase in the size of one breast” and/or “sticky or bloody discharge from one/both nipples”.	Examination by HSA doctor for psoriasis, dermatitis, lipoma and other skin lesions, as well as breast examination.

Table 2.1 continued…

PRIMARY OUTCOMES		
Domain	**Postal Questionnaire**	**Health Examination**
Neurological status	Self-reported sensory symptoms in the past month: “Have you suffered from difficulty recognising hot from cold water?” “Have you suffered from difficulty feeling pain, cuts or injuries?” “Have you suffered from numbness, asleep feeling, or prickling sensation in hands or arms?” “Have you suffered from numbness, asleep feeling, or prickling sensation in feet or legs?” “Have you suffered from unusual sensitivity or tenderness of your skin when clothes or bedclothes rub against you?” Self-reported motor symptoms in the past month: “Have you suffered from difficulty undoing buttons?” “Have you suffered from problems with tripping or your feet slapping while walking?” “Have you suffered from feeling unsteady walking on even ground?” “Have you suffered from feeling unsteady walking in the dark?” Self-reported previous physician diagnosis of multiple sclerosis and/or motor neurone disease. Self-report item on “increased sensitivity to smells or odours.”	Biothesiometry for peripheral vibration sense and “Sniffin’ Sticks” for olfaction, both administered by HSA doctor. L’Anthony Desaturated test for colour vision, administered by nurse.

Table 2.1 continued…

PRIMARY OUTCOMES		
Domain	**Postal Questionnaire**	**Health Examination**
Sexual function and reproductive health	*Males*. Self-completed 15-item International Index of Erectile Function. Also two separate self-report items asking about loss of interest in sex and/or problems with sexual functioning. *Females*. Self-completed Female Reproductive Questionnaire, with items for number of pregnancies, miscarriage, still birth and/or live birth and/or problems falling pregnant.	Not applicable.
Mental health	Self-reported previous physician diagnosis of “anxiety” and/or “depression” plus “Has anyone in your immediate family ever suffered from depression?” Self-reported experiences of “fatigue” in the past month. In terms of self-recorded medications, use of anxiolytic medications and anti-depressants were identified.	Administration of the Kessler 10-item scale, General Health Questionnaire 12-item scale, Composite International Diagnostic Interview (CIDI) for depression and anxiety and neurasthenia module.
Neuropsychological status	Subjective Memory Complaint Questionnaire (MAC-Q).	Comprehensive neuropsychological assessment by HSA psychologist, including: Rey 15-item test, Mini Mental Status Examination, and tests of executive functioning, psychomotor speed, attention/working memory, new learning/memory, and visuospatial abilities.

Table 2.1 continued…

SECONDARY OUTCOMES		
Domain	**Postal Questionnaire**	**Health Examination**
Hearing	Not applicable.	Air and Bone Audiometry conducted by nurse.
Visual acuity	Not applicable.	Snellen Chart Test conducted by nurse.
Cardiovascular disease risk	Not applicable.	Blood test for Homocysteine, C-Reactive protein and Apoliproprotein-E genotype.
Balance	Not applicable.	Functional Reach Test conducted by doctor.
GENERAL HEALTH OUTCOME MEASURES		
Domain	**Postal Questionnaire**	**Health Examination**
Adverse gastrointestinal symptoms	Not applicable.	Faecal Occult Blood Test kit provided to participants by nurse for self-completion.
Pulse rate and chest sounds	Not applicable.	Seated pulse, lying pulse, standing pulse (radial). Auscultation of chest and lungs for decreased breath sounds, bronchial breath sounds, wheezes, rubs and crackles.
Preventive health screening	Not applicable.	Blood test for glucose and cholesterol concentrations.

2.3.3.1 Storage of blood and immortalisation of lymphocytes

Due to the length of time that has elapsed between the DSRS programs and the present study, it was deemed unlikely that acute or short-term effects of the exposures would be able to be detected or assessed. Instead, the current study concentrated on chronic or long-term effects. Consideration was also given to the lag time or induction period between exposure and some outcomes. For some possible outcomes, e.g. cancer, the period between exposure and assessment may not be long enough for effects to be manifested, particularly for the more recent maintenance programs (i.e. Spray Seal DSRS). A literature review indicated that no current existing bio-marker (see Glossary) was appropriate for the current study, but that a number were under development. Therefore bio-markers of both exposure and outcome were proposed for inclusion. This was discussed at length by the study investigators and members of the Scientific Advisory Committee (SAC), and a decision was made to defer the tests but still to collect DNA, serum and cell samples for possible future testing, once relevant bio-markers were available.

Blood samples were centrifuged: plasma and red blood cells have been stored. The “buffy coat” has been isolated and lymphocytes have been immortalised by transfection with Epstein Barr virus, providing a source of DNA for use in future bio-marker assays without the inconvenience of multiple blood collections. Blood is to be stored for a period of 50 years, a reasonable period of time given that some participants are now in their mid 20s. Participants who consented to give blood could refuse to provide consent for the long-term storage of their blood sample. Nursing staff at Health Services Australia identified these participants using specially-developed Queensland Medical Laboratory (QML) blood collection forms.

2.3.4 Analysis

A detailed description of methods used to assess and compare the outcomes is given in Chapter 6 (Analysis). This section provides a summary of the overall analysis strategy for the study.

In general, the same pattern of analysis was used for each outcome.

- The primary analysis was to compare the entire exposed group – which included persons identified as participating in Program 1, Program 2, Wing or Spray Seal Programs – to each of the comparison groups.

- Secondary analyses included:
  - comparing Program 1 participants to comparison groups
  - comparing Program 2 participants to comparison groups*
  - creating a dose-response curve. This involved defining a mild, moderate and prolonged exposure group based on the length of time spent in DSRS activities. Seeing a graded increase in risk would support a causative role for DSRS exposures in that outcome.

In trying to establish a cause-and-effect relationship for DSRS activities and various outcomes, it is important to show that any ill-effects are not due to some other associated variables, termed "confounders" (see Glossary). These potential confounders may be factors such as age, education level, and lifestyle habits (e.g. smoking or alcohol intake). For example, if most DSRS workers were also smokers, it would be important to discern if any ill-health was due to DSRS or to cigarettes. A full description of potential confounders is found in Chapter 8.

2.4 Organisation and administration

The Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP) is a collaborative study conducted by researchers from The University of Newcastle Research Associates (TUNRA) and the Hunter Medical Research Institute. The study is administered by the Department of Veterans' Affairs on behalf of the Department of Defence, and also involves staff from Health Services Australia (HSA), Queensland Medical Laboratories (QML), and the Queensland Repatriation Transport Unit.

2.4.1 The University of Newcastle Research Associates

TUNRA Ltd was incorporated in 1969 as a company limited by guarantee, by the Council of The University of Newcastle, to provide a broadly-based research, consulting and specialist education service to industry and the community at large. TUNRA has access to the

* The current study could only investigate differences between exposed and comparison groups for individuals who indicated involvement in Program 1 and/or Program 2. The Wing DSRS and Spray Seal Programs contained fewer individuals, which decreased the study's statistical power to conduct separate investigations of these groups only. However, the findings from the current study may well apply to individuals involved in these Programs.

extensive facilities available through the University and is a registered training organisation (RTO) under the Australian Recognition Framework. Working under the auspices of TUNRA Ltd, the main project office for SHOAMP is located at The Centre for Clinical Epidemiology and Biostatistics (CCEB), School of Medical Practice and Population Health, Faculty of Health, The University of Newcastle. Members of the SHOAMP project team are:

Chief Investigators

- Associate Professor Catherine D'Este, DipAppSci (Medical Radiography), BMaths, DipEd, DipMedStats, PhD. Associate Professor in Biostatistics, Centre for Clinical Epidemiology & Biostatistics, University of Newcastle.
- Dr John Attia, BSc (Physiology), MSc (Clinical Epidemiology), MD, PhD (Medical Genetics), FRCP(C). Senior Lecturer in Epidemiology, Centre for Clinical Epidemiology & Biostatistics, University of Newcastle. Academic Consultant, Hunter Area Health Service.
- Dr Anthony Brown, MB BS, MPH, FAFPHM, FAFOM. Director, Population Health, Macquarie Area Health Service. Conjoint Associate Professor, Environmental and Occupational Health, University of Newcastle.
- Professor Julie Byles, BMed, PhD. Director, Centre for Research and Education in Ageing (CREA), Faculty of Health, University of Newcastle.

Associate Investigators

- Associate Professor Peter W. Schofield, BSc, MBBS, MSc (Epidemiology), MD, FRACP. Clinical Director of the Neuropsychiatry Service and Senior Staff Specialist in Neurology, Hunter Area Health Service. Conjoint Associate Professor of Psychiatry, University of Newcastle.
- Associate Professor Robert Gibberd, BSc (Hons), PhD. Associate Professor, Centre for Clinical Epidemiology & Biostatistics, University of Newcastle.
- Mr Steve Lee, BSc (Hons), MSc (Clinical Neuropsychology). Senior Clinical Neuropsychologist, Neuropsychiatry Service, Hunter Area Health Service.

CEO of TUNRA Ltd

- Dr Soozy Smith, PhD. TUNRA Ltd, University of Newcastle.

Project Management

- Ms Meredith Tavener, BAppSci (Hons), GradDip (Health Promotion), MMedSci. Project Manager, Centre for Clinical Epidemiology & Biostatistics, University of Newcastle.

Project Statistical

- Mr Richard Gibson, BSc, DipEd, DipMedStats. Associate Lecturer in Biostatistics (Research), Centre for Clinical Epidemiology & Biostatistics, University of Newcastle.

PhD Fellow

- Mrs Maya Guest, BOH&S, BMedSci (Comm. Health)(Hons). Research Higher Degree candidate, PhD Fellow for SHOAMP.

Information Management

- Mr Shane Jenkins, BA (Hons), MA (Inform. Tech.). IT and Data Manager.
- Ms Jan Mcleod, BA, DipEd. Information Manager.
- Ms Debbie Quain, RN, BA (Nursing). Research Assistant.
- Ms Marina Bernhard, BMaths. Statistical Assistant.
- Miss Rowena Brown. Research Assistant (literature review).
- Mr Ben Oastler. Research Assistant (data management).
- Mr Tim Moore, BEng (Comp)(Hons), BCompSci. Research Assistant (data management).

2.4.2 Hunter Medical Research Institute

The Hunter Medical Research Institute (HMRI) is a virtual institute that allows autonomous research groups located at different sites to be united by shared management structure, resources and philosophy. The SHOAMP team involved members of the HMRI with specific content or methodological expertise to provide the necessary high level of supervision to ensure outcomes of a high standard.

2.4.3 The Scientific Advisory Committee

The Scientific Advisory Committee (SAC) oversaw all scientific aspects of the Study of Health Outcomes in Aircraft Maintenance Personnel. This included work relating to:

- the literature review
- protocols for qualitative interviews
- development of the Protocol Manual for the General Health and Medical Study
- overseeing which specific health tests were to be administered as part of the General Health and Medical Study
- which items were to be included in the Postal Questionnaire
- how data were to be returned to the research team
- which analyses were to be conducted.

The SAC standing members were:

- Professor Judith Whitworth AC (SAC Chair), MD, BS, DSc, PhD, (Melb), FRACP, FAICD. Director, The John Curtin School of Medical Research, Australian National University.
- Professor Michael Moore, BSc (Hons)(Biochemistry), PhD, DSc (Biochemistry in Medicine). Director NH&MRC National Research Centre for Environmental Toxicology, University of Queensland.
- Professor Bruce Armstrong AM, BMedSc (Hons), MBBS (Hons), DPhil, FRACP, FAFPHM, FAA. Head, School of Public Health, University of Sydney. Honorary Epidemiologist, The Cancer Council NSW. (Predecessor to Professor Roder; attended the SAC from 7-11-2000 to 28-5-2001.)
- Professor David Roder, DDSc, MPH, AM. Consultant Epidemiologist, The Cancer Council South Australia. (Successor to Professor Armstrong, attended the SAC from 6-11-2001.)
- Dr Deborah Glass, MA (Cantab), MSc, PhD, DipOccHyg. Department of Epidemiology and Preventative Medicine, Monash University. Occupational Hygiene Unit, School of Chemical and Biological Sciences, Deakin University.
- Emeritus Professor Scott Henderson AO, MD (Aberd), Hon MD (UNSW), DSc (ANU), FRCP, FRACP, FRCPsych, FRANZCP. Visiting Fellow, The John Curtin School of Medical Research, Australian National University.

In addition to standing members, the SAC is also regularly attended by representatives from a number of other organisations:

- Department of Veterans' Affairs
- Defence Health Services Branch
- Air Force Headquarters
- F-111 Advocate's Office
- Australian Institute of Health and Welfare
- Health Services Australia
- SHOAMP Consultative Forum.

2.4.4 The Consultative Forum

A Consultative Forum provided a link between the SAC and interested parties. The Forum received regular briefings on proposals in relation to the conduct of the study and provided an opportunity for feedback from members on issues such as privacy, storage of information, and selection of comparison groups. Organisations represented at the Consultative Forum included:

- Department of Veterans' Affairs
- TUNRA study team
- Defence Health Services Branch
- Air Force Headquarters
- F-111 Advocate's Office
- SHOAMP SAC
- Warrant Officer of the Air Force
- Australian Veterans and Defence Services Council
- Armed Forces Federation of Australia
- Regular Defence Force Welfare Association
- Royal Australian Air Force Association
- Defence Community Organisation
- Returned Services League of Australia Limited (RSL)
- F-111 DSRS Support Group
- F-111 DSRS Partner's Support Group
- SERCO Defence Services
- Repatriation Medical Authority
- Queensland Workcover

- Representatives from Deseal / Reseal programs
- Health Services Australia.

2.4.5 DVA Contact and Recruitment Team

The Department of Veterans' Affairs (DVA) provided a team to coordinate contacting and recruiting participants for SHOAMP. The role of the Contact and Recruitment (C&R) Team was to put together and mail out study invitation documentation to potential participants from both the exposed and comparison cohorts, and then to apply follow-up procedures for non-respondents where necessary. The C&R Team was also responsible for posting out reminder cards and replacement invitation packages to non-respondents. Once contact had been established with a participant, the Team sought clarification as to which aspects of the study the participant was consenting to participate in (e.g. completion of the Postal Questionnaire but not attending a health examination). If participants agreed to a health examination, then an appointment was made directly with Health Services Australia (HSA). The C&R Team (see Glossary) also facilitated transport arrangements for participants (together with the Queensland Repatriation Transport Unit) and processed claims for loss of earnings. Personnel included:

- Mr Arthur Edgar
- Ms Heather Parry
- Ms Peta Stevenson
- Dr Keith Horsley
- Dr Warren Harrex
- Dr Eileen Wilson
- Mr David Goldrick
- Mr Tim Beard
- Mr Ces White
- Mr Barry Miles
- Mr David Steer.

2.4.6 Queensland Repatriation Transport Unit

The Queensland Repatriation Transport Unit (RTU) provides transport assistance to the eligible veteran community to access their required medical treatment. The Queensland RTU provided transport assistance to the study by arranging and booking air, train, bus and taxi

journeys throughout Australia. The RTU also processed claims by study participants for travel reimbursements, including the reimbursement of petrol, meals and accommodation costs.

2.4.7 Health Services Australia

Operating from over 13 offices in rural and urban Australia, Health Services Australia (HSA) is the most widespread provider of workplace health and safety medical assessment services in the country. HSA worked in collaboration with DVA and the TUNRA research team to conduct health examinations for consenting SHOAMP participants. Key participants in the SHOAMP study were:

- Mr Stan Macionis, BE(Hons), MBA, FAIM. General Manager, HSA Queensland.
- Dr Carol Toft, MBBS(Hons). Senior Medical Advisor, HSA Queensland.
- Ms Rhonda Cameron. Business Development Officer, HSA Canberra.

2.4.8 Queensland Medical Laboratories

Working closely with Health Services Australia, Queensland Medical Laboratories (QML) provided a collection and processing service for SHOAMP blood samples. QML provided all blood collection tubes, de-identified* instruction forms for blood processing, and worked with World Courier to have all samples transported to their laboratories on the same day as collection.

* Technical term meaning that there was no identifying data on the form that could be linked to someone's identity.

This page has been left blank intentionally
for formatting purposes

3 Study Population and Recruitment

Chapter summary

In the absence of any definitive record of F-111 Deseal/Reseal (DSRS) involvement, the "exposed" group was selected through various means, including those identified by the Board of Inquiry (BOI), advertisements in newspapers and relevant newsletters, a free-call hotline and direct enquiries. Contemporaneous comparison groups from Amberley (same Base, different job) and Richmond (different Base, similar job) were identified from personnel files. Contact was made by post, telephone or email. Participation rates were good for the exposed group (77%) but low for the comparison groups (40% for Amberley, and 48% for Richmond) despite multiple attempts and routes of follow-up, a disparity which leads to potential selection bias. Nevertheless, characteristics such as rank, posting and age were reasonably well-matched across participant groups.

Chapter contents

3.1 Introduction

Objectively identifying the “exposed” group (i.e. all personnel who participated in any F-111 Deseal/Reseal activities since 1975) proved to be a particularly difficult task. There was no system of direct notations on personnel files for those who were involved with DSRS and hence no official record of those involved. Several roundabout ways of identifying the group of interest were considered, including reviewing aircraft log books, tracing extra pay given to those who entered the tanks, and identifying those who took a “confined spaces entry” course. However, none of these approaches proved feasible.

3.2 General approach

The general contact and recruitment strategy for the Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP) involved several processes. Eligible personnel from exposed and comparison groups were mailed an invitation and information package. The documentation in this package encouraged recipients to call the Department of Veterans’ Affairs (DVA) directly, using a 1800 free-call number, to register themselves for all, or some parts only, of the study, or to refuse to take part. Those who did not respond to the first invitation within one week of the initial mail-out were sent a reminder card, which prompted them to register their interest with DVA. This card also provided contact details for the study team, the Airman’s Advocate, and representatives of the Ethics Committee. Eligible personnel who did not respond to the reminder card within one week (i.e. within two weeks of initial mail-out) were telephoned (if their telephone contact details were available) by a member of the DVA Contact & Recruitment (C&R) Team to seek their decision regarding participation.

All invitees were given the opportunity to call a second 1800 free-call number to have their questions answered by the SHOAMP Project Manager. They were also given contact numbers for an independent representative from one of three participating Ethics Committees (The University of Newcastle Human Research Ethics Committee, The Australian Defence Force Human Research Ethics Committee, and the Department of Veterans’ Affairs Ethics Committee).

For invitees who gave their consent, the C&R Team made an appointment for a SHOAMP health examination through a direct electronic link with the Health Services Australia (HSA) appointment diary. The C&R Team worked with HSA to post out a letter to consenting participants to confirm their health examination appointment. Also included with this letter were participant instructions and pro formas for claiming travel and loss-of-income costs. Those who did not wish to participate in the health and medical examination were asked to complete the Postal Questionnaire and return it to the research team using a reply-paid envelope included in the mail-out package.

Written informed consent was obtained from all participants who chose to return their Postal Questionnaire and/or take part in a health examination in collaboration with HSA. Participants were able to withdraw from the study at any time.

The initial mail-out package consisted of documents that described the Health Study, outlined what would be expected of participants, and gave reassurance as to the privacy of any information provided. Potential participants from the exposed group and comparison groups were sent the same information, except that those in the comparison groups were *not* sent the questionnaire that asked them to clarify their F-111 DSRS experiences (the Exposure Questionnaire).

The initial mail-out package included the following:

- Letter of support from the Minister for Defence (Appendix 3A)
- Letter of support from the Consultative Forum (Appendix 3B)
- Information and Invitation Statement (Appendix 3C)
- Informed Consent Statement (Appendix 3D)
- Postal Questionnaire – Male Version (Appendix 2K) *or* Postal Questionnaire – Female Version (Appendix 2L, different items only)
- Exposure Questionnaire, sent to exposed category “1” personnel only (Appendix 2I)
- Female Reproductive Questionnaire, for the female partners of male study participants (Appendix 2M)
- Australian Defence Force Human Research Ethics Committee “guidelines to volunteers” (Appendix 3E)
- Reminder card, for all non-respondents (Appendix 3F).

Confirmation-of-appointment documentation included:

- Letter confirming the HSA health examination appointment (Appendix 3G)
- Guidelines to health examination participants (Appendix 3H)
- Travel Claim Form (Appendix 3I)
- Loss of Earnings Claim Form (Appendix 3J).

3.3 Sampling strategy

The starting point for creating a sampling frame was taken from the list of DSRS individuals identified by the Board of Inquiry (BOI) and expanded over time as additional people contacted DVA in response to media advertising and word of mouth. The development of the DVA F-111 DSRS list was described in Chapter 2 (Overview of the General Health and Medical Study).

The list maintained by DVA included a variety of data relating to individuals' self-reported involvement in DSRS activities, such as time of exposure, type of exposure (e.g. Program 1, Program 2) and rank at time of exposure, in addition to information on age, date of birth, and problems experienced as a result of involvement with DSRS. DVA personnel loosely categorised each respondent's level of exposure to DSRS tasks into three groups, defined in Chapter 2:

- *Exposure Category 1* – at any time directly involved with the Amberley F-111 Deseal/Reseal programs
- *Exposure Category 2* – not directly involved with the DSRS programs but worked in such proximity as to be at risk
- *Exposure Category 3* – at RAAF Base Amberley during the exposure period of interest (i.e. conduct of the DSRS programs) but not directly involved and did not work within such proximity as to be at risk.

This process continued until the database of participants was "locked down" on 12 September 2002 so that health examinations could begin. A small number of people continued to identify themselves to DVA after this date; and an extension was given so that all those reporting up to the end of March 2003 could potentially be included in the Health Study. Anyone who contacted DVA after the end of March 2003 had their details recorded but were not included in the health study. Individuals classified by DVA as category 2 or 3

and those registered on the database prior to March 2003 were mailed an Exposure Questionnaire (EQ) to determine their eligibility for the Health Study.

Data for the comparison group sampling frames were provided by DVA from the computerised Air Force Personnel Executive Management System (AFPEMS). Two different files were provided. The first was a file of personal details, and included service number, gender, date of birth, date of enlistment, and date of death (if applicable), with one record for each individual. The second file included posting details, with one record for each posting or attachment to the relevant RAAF Base within the study time period. Separate files were provided for Amberley and Richmond comparisons. The AFPEMS files included all individuals who had been posted in to RAAF Base Amberley with a non-technical mustering (for comparison group 1) or RAAF Base Richmond with a technical mustering (for comparison group 2) between 1 January 1975 and 31 December 1999. A brief description of the procedure for obtaining the comparison groups is provided in Appendices 2E and 2F.

Using the AFPEMS database, some 10,100 personnel were listed as having a total of 34,322 new postings or attachments for duty to RAAF Richmond between 1975 and 1999. Similarly, some 11,164 individual personnel were identified as having had 48,644 postings or attachments to RAAF Amberley over the same period. After excluding individuals known to have died, multiple records for individuals within each posting period, records which did not have the appropriate combination of matching characteristics, and records for individuals identified as being in the exposed group, there were in total 13,549 and 11,037 eligible postings for the Richmond and Amberley comparison groups respectively. An individual could be an eligible comparison in more than one posting period.

For each exposed individual, two potential comparisons (one in each comparison group) were selected with the same combination of gender, five-year age group, posting category and rank category. Duplicate records for individuals were then excluded. Individuals who were selected for both Amberley and Richmond comparison groups were included in the Amberley comparison group (as this was the first group selected) and excluded from the Richmond comparison.

The Contact and Recruitment Team (C&R Team) at the Department of Veterans' Affairs (DVA) was responsible for: mailing out all invitation packages to both the exposed and comparison groups; recording contact details and consent status; conducting follow-up telephone calls; and arranging appointments for participants' health examinations using the

computerised Electronic Management of Medical Assessments (ELMA) diary system at HSA. Recruitment was conducted in three main phases:

1) *Round One* – a pilot phase, during which all contact and health examination processes were trialled.
2) *Round Two* – the first phase of the main body of data collection for the General Health and Medical Study.
3) *Round Three* – the second and final phase of data collection for the General Health and Medical Study, during which individuals were included who had a DVA classification of "2" or "3", or who registered later with the F-111 DSRS Hotline and were determined to be eligible (based on their responses on the Exposure Questionnaire).

3.4 Round One

Round One was conducted in one HSA centre only. Brisbane was the selected site because it was anticipated that the largest proportion of participants would reside in the Brisbane area. A selection strategy for Round One was developed, based upon the following steps, to identify exposed personnel who were to receive an invitation package:

1) Identification of personnel from the F-111 DSRS database who had a classification of "exposed", meaning that they had definitely been involved in DSRS (n=719).
2) Priority identification of individuals within the Brisbane area postcode of 4000-4199 (n=101).
3) From the priority sample of eligible participants, females, due to their smaller number, were purposefully selected to ensure that they received an invitation to take part so that female-based questionnaire items could be trialled (n=2).
4) From the remaining eligible male participants (n=99), observations were sorted by DVA F-111 list ID number from which every 5th individual was selected (n=19).
5) As no civilians were identified through random sorting and selection, all eligible civilians were then included (n=4).

This resulted in a total sample of 25 "exposed" who were to receive a full invitation package. It was agreed that a sample of 50 comparisons would be identified, assuming a lower contact and consent rate for this group relative to the exposed group. The comparison individuals

were randomly selected from the file of Amberley non-technical postings and were frequency-matched to the exposed group on the following variables:

- gender
- age (5-year increments starting at less than or equal to 19, 20-24, 25-29, 30-34 etc.)
- date of involvement in the program (or posting) by year groups (1975-79, 80-84, 85-89, 90-94, 95-1999)
- rank at posting or time of DSRS involvement (for exposed).

Amberley comparisons only were considered for Round One because the Round One health examinations were to be conducted only at Brisbane HSA. It was initially anticipated that most of the Amberley comparison group would reside in Southern Queensland and would therefore be eligible for Round One.

Mustering (Defence Force job categorisation) had initially been considered as a stratification variable in addition to those above. However, this was not used for two reasons. Firstly, including mustering as an additional matching variable would have resulted in a very large number of possible strata and potential difficulty in finding appropriate matches. Secondly, the comparison groups were initially defined by mustering: technical trades for the Richmond comparison and non-technical trades for the Amberley comparison. The mustering categories for each comparison group have been detailed in Appendices 2B and 2C.

One of the 25 exposed individuals selected for Round One was later found to reside outside the Brisbane area and was therefore excluded. This individual was reclassified as eligible for Round Two. Only six of the 50 selected comparison individuals lived within the Brisbane area, and a seventh person was discovered to have died. This meant it was necessary to select further comparisons for Round One: 41comparisons were selected.

Invitation packages were therefore posted out to 24 exposed and 47 comparison individuals (six of the initial Round One comparisons and 41 of the additional comparisons) on 1 October 2002. After seven days, a reminder card was posted out to non-respondents; and after another seven days, the DVA C&R Team began to follow-up non-respondents by telephone. Training of HSA personnel was conducted for Round One on 14 October 2002, and the first health examinations were conducted on 16 October 2002.

From the 71 eligible participants invited to take part in Round One (24 exposed and 47 comparison), a total of 30 individuals received a health examination during the period 16 October to 18 November 2002 (14 in the exposed group and 16 in the Amberley comparison

group). This constituted consent rates of 58% and 34% for exposed and comparison groups respectively. Some individuals who could not attend a health examination during the Round One study period participated in Round Two.

3.5 Round Two

3.5.1 Exposure Category 1 individuals

Those individuals selected to participate in Round One who did not take part in a health examination were eligible to be assessed in Round Two. Members of the exposed group who were eligible to participate in Round Two were identified from the DVA F-111 database as having the classification of “1” (definitely involved in F-111 DSRS). Individuals in categories “2” and “3” were reviewed at a later date for possible reclassification as exposed (see Section 3.6 Round Three).

Prior to the Round Two mail-out, DVA used Defence and DVA records to update any incomplete contact details for potential participants in order to facilitate a successful mail-out process. Mail-out to potential participants was staggered to allow for different start and finish dates for the various HSA centres. The first mail-out was to individuals whose DVA records indicated that they lived within the Brisbane and Ipswich postcode areas. This strategy minimised the amount of time between Rounds 1 and 2 for the HSA personnel in Brisbane and Ipswich, and hence minimised the need for them to undergo “refresher” training.

In total, 3163 personnel were selected to take part in Round Two. This figure comprised 695 individuals from the F-111 database classified as exposure level “1”, 1204 Amberley comparisons (including those initially selected for Round One but residing outside the Brisbane area), and 1264 Richmond comparisons. There was a total of 2468 eligible comparison participants.

As with Round One, training for HSA professionals was conducted as close as possible to the commencement of Round Two health examinations. However, due to equipment requirements and staffing resources between centres, commencement dates had to be staggered.

3.5.2 Overseas participants

As well as Australian addresses, several people in DVA's F-111 database had overseas contact details (which also changed according to deployments over time). It was not uncommon for aircraft personnel to continue to work overseas in the field of aircraft maintenance following their discharge from the Australian Defence Force, one of the most common places being the Saudi Arabian Airforce.

Individuals from both the exposed and comparison groups were sent an invitation package for the health study. Since the 1800 free-call number would not work outside Australia, slight modifications were made to the mail-out materials: recipients were asked to provide fax, email and postal contact details; and a consent form was included that they could fax or email back to the study team or DVA. The Postal Questionnaire could be posted back to the study team, and all efforts were made by the C&R Team at DVA to coincide health examination appointments with participants' return dates to Australia. Seven overseas participants returned a Postal Questionnaire: three in the exposed group, and four in the Richmond comparison group. Two overseas individuals in the exposed group, three in the Amberley comparison group and four in the Richmond comparison group refused to participate in all parts of the Health Study.

3.6 Round Three

3.6.1 Exposure Category 2 and 3 individuals

In addition to the category 1 exposed group (those definitely involved in DSRS activities) and the Amberley and Richmond comparison groups, there was a third group of individuals identified from DVA's F-111 DSRS database. These individuals were classified as category 2 (worked in close proximity to DSRS activities) or category 3 (at Amberley RAAF Base during the exposure period of interest, but not involved directly with DSRS activities). These categories indicated some potential exposure to DSRS activities but more information was needed before a final decision could be made. A total of 500 individuals were in categories 2 and 3, and each individual was sent an Exposure Questionnaire during the Round One and Round Two mail-outs (see Section 3.6.3 below). This mail-out occurred after all exposed and

comparison invitation packages had been posted out. Based on the 283 (57%) returned questionnaires, 143 individuals (51%) were reclassified as exposed.

3.6.2 Ongoing F-111 database registrations

Following the first lockdown of the DVA F-111 database in September 2002, new callers were still able to contact DVA and leave details of their F-111 DSRS work experiences. New callers had their details recorded by the DVA C&R Team, and each person was sent out an Exposure Questionnaire. A second lockdown date was determined (31 March 2003) after which new callers continued to have their personal details recorded by DVA but were not included as part of the study population. A total of 146 people registered with DVA between the first and second lockdown dates (one of whom died before study recruitment). A total of 259 registrations were received by DVA after the second lockdown. Callers who registered between the first and second lockdown dates (n=146) were posted out an Exposure Questionnaire (see Section 3.6.3 below). As a result of this mail-out, 109 individuals returned a questionnaire (75%), 47 (43%) of whom were classified as exposed.

3.6.3 Reclassification

As mentioned previously, the classification of exposure by DVA was only semi-structured. In order to be consistent and rigorous, a more detailed assessment of exposure was required. Exposure Questionnaire (EQ) data were used as the basis for this reclassification into “exposed” or “non-exposed”.

The EQ had first-level screening items which asked respondents to indicate if they were involved in any of the DSRS programs (Program 1, Program 2, Wing Program or Spray Seal Program) or in the activities of storing, mixing and/or disposing of DSRS-related chemicals. Where a person gave a positive response to a first-level item, they were then asked to provide further detail of their work activities using second-level questions which described a variety of work duties known to have been involved in the different program types. Third- and fourth-level EQ responses were to questions about the amount of time spent on each nominated DSRS activity. Each survey was reviewed both electronically and, if comments had been written in the free text fields, manually. A respondent was considered to be “exposed” if they gave a “yes” response to any first-, second-, third- or fourth-level screening item, or if they provided comments that gave the same indication of involvement as a first-level screening item. All “new exposed” individuals were reclassified accordingly and the list

of ID numbers communicated to DVA so that the full Health Study invitation mail-out package could be sent out. Electronic and manual methods for determining exposure are discussed in detail in Chapter 4.

Of the 719 individuals initially classified as exposed by DVA, 491 (68%) returned an Exposure Questionnaire. Thirty-seven of these were then re-classified as unexposed and excluded from the Health Study. The selection of comparison individuals was based on the characteristics of DVA category 1 (Round One and Two) exposed individuals. After reclassification of DVA category 2, 3 and new registrations, characteristics of those individuals deemed to be eligible for the Health Study were compared to the Amberley and Richmond comparison groups on gender, age category, posting category and rank category to determine whether it was necessary to select any further comparisons. As the two comparison groups and two exposed groups (previously and newly exposed) were similar, no additional comparisons were selected.

3.7 Measures to maximise contact and follow up

As part of the overall contact and recruitment process, DVA implemented an ongoing checking process to update the contact details of eligible study participants where necessary. These checks identified a large number of eligible SHOAMP participants, from both the exposed and comparison groups, whose records did not contain up-to-date address and telephone contact details. In order to boost contact with potential study participants, DVA requested that SHOAMP data be matched against personal information held by the Australian Electoral Commission (AEC) and the Health Insurance Commission (HIC). In addition to this, regular checks were made of personal contact details against the online Telstra White Pages facility. The inclusion of data-matching exercises with agencies external to the Department of Veterans' Affairs and the Department of Defence complies with privacy principles and is governed by paragraph 130(3)(a) of *the Health Insurance Act 1973*.

3.7.1 Australian Electoral Commission

The full list of exposed and comparison individuals was sent to the Australian Electoral Commission (AEC) prior to Round One (September 2002) for matching against address and telephone details. Contact information was matched for approximately 80% of exposed

individuals and 81% of comparison individuals. A second round of matching was conducted in February 2003 when the final exposed and comparison groups were submitted once more to the AEC, again resulting in an 80% match.

3.7.2 Health Insurance Commission

A similar request to match up-to-date addresses and telephone numbers was also sent to the HIC in February 2003. Of the total number of potential participants, 90% were returned with exact matches of surname, first name, initial, and date of birth; and 1.5% were found to be deceased.

3.7.3 Telstra Electronic White Pages

The Telstra Electronic White Pages were used by DVA to identify changes of address for participants who had “return-to-sender” on mailed documentation and/or could not be contacted using their original address details. The White Pages provide up-to-date details of change of address on the Internet within a seven-day period of issue. The C&R Team re-checked the White Pages information on a regular basis, seeking a match between addresses and potential participants to further the contact process.

3.7.4 Introduction of “Flyer”

In addition to the original mail-out materials developed for potential study participants, a “flyer” (Appendix 3K) was developed to further encourage their participation. The flyer was posted out to the following classifications of potential participant:

- *Unknown* – those who had previously been sent an invitation package and/or had been contacted by telephone and remained indecisive about any form of participation; those who had not yet been contacted by telephone in person but who had been left a message; and those who had only been sent a postal package but who were yet to be contacted by telephone.
- *Non-contactable* – those who had been posted a study invitation package but had not responded and did not have telephone contact details listed on the F-111 DSRS database.

3.7.5 Defence superannuation

The Defence Superannuation administration was approached and asked to:

1) provide contact details for the selected individuals
2) confirm that the contact details recorded in the DVA contact and recruitment database were accurate
3) forward a recruitment package to individuals on behalf of DVA and TUNRA.

Defence Superannuation declined to do any of these three things because they believed it contravened privacy policy. The issue was not taken further.

3.7.6 Defence email

Over 300 “blind” emails (i.e. name@defence.gov.au) were sent by the C&R Team in an effort to contact serving members for whom participation status in the study had not been confirmed. This resulted in two participants who agreed to the Postal Questionnaire and health examination; two who agreed to the questionnaire only; and two who refused.

3.7.7 Timing of follow-up phone calls

Where an individual’s participation status for the Health Study was still unknown, the C&R Team made up to ten attempts to call them. The calls were made after working hours and also on weekends. If callers got through to an answering machine, they left a message in the hope that the person would be in contact. If callers were switched through to a fax machine, they then faxed a letter on Departmental letterhead asking the person to contact the C&R Team to confirm the phone number.

3.8 Final summary of recruitment and participation

The final figures for participation (combining recruitment Rounds One, Two and Three) are shown in Table 3.1. Round Two and Round Three recruitment is summarised in Appendix 3L.

Table 3.1 : Final participation figures for the General Health and Medical Study

Identified	Total number of participants		
	Amb	Rich	Exp
Refused	420	373	155
Deceased	20	23	14
Non-contactable	316	268	44
Contacted and consented			
Full participation	400	508	592
Health examination only (full or part)	6	8	24
Postal questionnaire only	89	84	43
Totals	1251	1264	872

A total of 872 exposed individuals (24 from Round One, 695 from Round Two and 190 from Round Three less 37 classified as ineligible), 1251 Amberley comparisons (47 from Round One and 1204 from Round Two) and 1264 Richmond comparisons were eligible for inclusion in the Health Study. Of these, 57 were known to have died (20 in the Amberley comparison group, 23 in the Richmond comparison group and 14 in the exposed group). These deaths were excluded from the denominator in the calculation of response rates. In total there were 659 individuals in the exposed group who participated to some extent in the Health Study, giving a final participation rate of 77%. Of these, 592 participated in the full study (90%), 24 in the health examination only (4%), and 43 returned the Postal Questionnaire but did not have a health examination (6%).

There were 400 individuals in the Amberley group who participated in the full study, six who only completed a health examination, and 89 who only returned a Postal Questionnaire. Overall 40% of the Amberley comparison group participated in some component of the General Health and Medical Study. Of the Richmond comparison group, 508 individuals

participated in the full study, eight only completed a health examination, and 84 only returned a Postal Questionnaire, giving a response rate of 48% for any component of the study. A large proportion of the sample could not be contacted: 26% of the Amberley comparison group and 22% of the Richmond comparison group, but only 5% of the exposed group. Excluding these individuals from the denominator provides consent rates of 54%, 62% and 81% for the Amberley comparison, Richmond comparison and exposed groups respectively.

3.9 Exploration of group characteristics

Because additional exposed individuals were included in the study, and because participation rates from the two comparison groups were poor, it was necessary to explore whether characteristics of the exposed group were similar to those of the comparison groups and to examine whether respondents were different from non-respondents. This was explored in three ways:

1) Gender, age, posting, and rank categories were compared between all **eligible** exposed and **eligible** Amberley and Richmond comparison groups (shown in Section 3.9.1).
2) Gender, age, posting, and rank categories were compared between respondents and non-respondents within each of the three groups (shown in Section 3.9.2).
3) Gender, age, posting, and rank categories were compared between all **actual** respondents from the exposed and Amberley and Richmond comparison groups (shown in Section 3.9.3).

The “chi-square” statistic (see Glossary) was used for all comparisons. For age comparisons, the two youngest age categories and two oldest age categories were combined because of small numbers (although the original numbers are presented in the tables). The “civilian” and “unknown” rank categories were excluded from the chi-square analyses for between-group comparisons, because only the exposed group included civilians or individuals with missing rank.

3.9.1 Selection of the original sample

Table 3.2 shows the distribution of gender, age, posting and rank by group for all 3387 individuals **eligible** for the General Health and Medical Study. The three groups had similar

distributions for gender ($\chi_1^2 = 2.51, p = 0.29$), age ($\chi_{16}^2 = 6.75, p = 0.98$), posting ($\chi_8^2 = 5.05, p = 0.75$) and rank ($\chi_4^2 = 1.23, p = 0.87$).

Table 3.2 : Characteristics of eligible DSRS exposed and RAAF Amberley and Richmond comparison groups

	Amberley N=1251		Richmond n=1264		Exposed n=872	
	n	%	n	%	n	%
Gender						
Female	17	1.4	16	1.3	18	2.1
Male	1234	99	1248	99	854	98
Age Category						
≤ 24yrs	2	0.2	2	0.2	1	0.1
25-29yrs	64	5.1	66	5.2	41	4.8
30-34yrs	167	13	167	13	115	13
35-39yrs	233	19	237	19	157	18
40-44yrs	280	22	275	22	174	20
45-49yrs	259	21	263	21	170	20
50-54yrs	124	10	130	10	96	11
55-59yrs	70	5.6	72	5.7	63	7.3
60-64yrs	27	2.2	26	2.1	23	2.7
65-69yrs	20	1.6	20	1.6	16	1.9
70-74yrs	5	0.4	6	0.5	4	0.5
Posting Category						
1975-1979	397	32	415	33	295	34
1980-1984	292	23	289	23	186	22
1985-1989	300	24	300	24	194	22
1990-1994	160	13	153	12	125	14
1995-1999	102	8.2	107	8.5	65	7.5
Rank Category						
Civilian	0	0	0	0	48	5.5
Enlisted	836	67	843	67	515	59
Non-Comm. Officer	385	31	391	31	246	28
Officer	30	2.4	30	2.4	24	2.8
Unknown	0	0	0	0	39	4.5

Includes 57 individuals known to have died.

3.9.2 Response bias

Postal Questionnaire

Table 3.3 shows the distribution of gender, age, posting and rank categories for Postal Questionnaire respondents and non-respondents by group. There was no difference in gender for respondents and non-respondents in any of the three groups, although this comparison almost reached statistical significance for the Amberley comparison group ($\chi_1^2 = 3.51, p = 0.06$; $\chi_1^2 = 0.68, p = 0.41$; $\chi_1^2 = 0.83, p = 0.36$ for comparison of gender by response rates for Amberley, Richmond and exposed groups respectively). Age distribution was similar for those who did and did not complete a Postal Questionnaire for the Amberley comparison ($\chi_8^2 = 13.80, p = 0.09$) and exposed groups ($\chi_8^2 = 7.96, p = 0.44$). For the Richmond comparison group, there were higher proportions of respondents relative to non-respondents in the older age groups (40 years and over) ($\chi_8^2 = 56.40, p < 0.0001$).

Table 3.3 : Comparison of characteristics of Postal Questionnaire respondents and non-respondents by group

	Amberley		Richmond		Exposed	
	Respondents N=489 %	Non-Respondents N=742 %	Respondents N=592 %	Non-Respondents N=649 %	Respondents N=635 %	Non-Respondents N=223 %
Gender						
Female	0.6	1.9	1.0	1.5	2.4	1.4
Male	99	98	99	98	98	99
Age category						
20-24yrs	0	0.3	0.2	0.2	0.2	0
25-29yrs	4.1	5.9	3.4	7.1	4.9	4.7
30-34yrs	11	15	9.0	18	13	15
35-39yrs	19	19	15	22	17	23
40-44yrs	25	21	26	19	20	21
45-49yrs	20	21	23	19	20	18
50-54yrs	10	9.3	13	7.7	12	9.4
55-59yrs	7.0	4.3	7.1	4.0	7.9	5.2
60-64yrs	2.0	2.0	1.7	2.2	2.7	1.4
65-69yrs	1.2	1.5	1.2	0.9	1.7	2.4
70-74yrs	0	0.3	0.3	0.5	0.5	0.5
Posting category						
1975-1979	32	30	38	27	35	31
1980-1984	24	23	24	22	21	22
1985-1989	25	24	23	25	21	27
1990-1994	12	14	9.3	15	15	14
1995-1999	8.0	8.4	5.7	11	7.9	6.9
Rank category						
Civilian	0	0	0	0	6.1	4.0
Enlisted	62	71	62	73	58	63
Non-Comm. Officers	34	2.7	35	25	30	23
Officers	3.3	1.9	2.9	1.9	3.0	2.2
Unknown	0	0	0	0	3.3	7.6

Excludes individuals known to have died.

Posting categories were also similar for respondents and non-respondents for the Amberley Comparison group and the exposed group ($\chi^2_4 = 1.33, p = 0.86$ and $\chi^2_4 = 3.71, p = 0.45$ respectively), but differed for the Richmond comparison group ($\chi^2_4 = 32.45, p < 0.0001$). Respondents had more individuals in the earliest period (38%) than non-respondents (27%). Within all three groups, there were more Postal Questionnaire respondents than non-respondents with higher rank categories ($\chi^2_2 = 10.18, p = 0.006$ for Amberley and $\chi^2_2 = 16.86, p = 0.0002$ for Richmond and $\chi^2_2 = 12.05, p = 0.017$ for exposed).

Health examination

Health examination participants were more likely to be male than non-participants in the Amberley comparison group ($\chi^2_1 = 5.73, p = 0.017$), but gender was evenly distributed for participants and non-participants in the Richmond comparison group ($\chi^2_1 = 0.71, p = 0.40$) and the exposed group ($\chi^2_1 = 1.21, p = 0.27$) (see Table 3.4).

Table 3.4 : Comparison of characteristics of health examination participants and non-participants by group

	Amberley		Richmond		Exposed	
	Participants N=406 %	Non-Participants N=825 %	Participants N=516 %	Non-Participants N=725 %	Participants N=616 %	Non-Participants N=242 %
Gender						
Female	0.3	1.9	1.0	1.5	2.4	1.2
Male	100	98	99	98	98	99
Age category						
20-25yrs	0	0.2	0	0.3	0.2	0
25-29yrs	3.5	6.1	3.3	6.8	5.2	3.9
30-34yrs	10	15	9.3	16	13	15
35-39yrs	19	19	15	22	17	21
40-44yrs	26	21	26	19	20	21
45-49yrs	21	20	24	18	20	19
50-54yrs	10	9.5	13	8.1	11	11
55-59yrs	7.4	4.4	6.8	4.6	8.1	4.8
60-64yrs	1.5	2.3	1.4	2.3	2.9	0.9
65-69yrs	0.7	1.7	1.0	1.1	1.8	2.2
70-74yrs	0	0.2	0.4	0.4	0.5	0.4
Posting category						
1975-1979	33	30	37	29	35	31
1980-1984	25	23	25	21	21	22
1985-1989	24	25	23	24	21	27
1990-1994	11	14	9.1	14	15	13
1995-1999	6.9	8.9	5.2	11	7.5	8.1
Rank category						
Civilian	0	0	0	0	6.2	4.1
Enlisted	64	69	62	71	58	62
Non Comm. Officers	33	29	35	27	29	25
Officers	3.0	2.2	2.9	1.9	2.9	2.5
Unknown	0	0	0	0	3.6	6.6

Excludes individuals known to have died.

For both Amberley and Richmond comparison groups there were generally more participants than non-participants in the over 40 years of age categories ($\chi^2_8 = 21.36, p = 0.006$ for Amberley and $\chi^2_8 = 50.97, p < 0.0001$ for Richmond), while age was similar for participants and non-participants in the exposed group ($\chi^2_8 = 8.35, p = 0.40$). The distribution of posting categories was similar for health examination participants and non-participants in the Amberley comparison and exposed groups ($\chi^2_4 = 3.94, p = 0.41$, $\chi^2_4 = 4.56, p = 0.34$ respectively), but participants in the Richmond group had earlier posting categories than non-participants ($\chi^2_4 = 26.78, p < 0.0001$). For the Amberley comparison group and the exposed group, rank categories were similarly distributed for participants and non-participants ($\chi^2_2 = 3.80, p = 0.15$ for Amberley and $\chi^2_2 = 6.57, p = 0.16$ for exposed). Within the Richmond comparison group, participants had lower ranks than non-participants ($\chi^2_2 = 10.74, p = 0.005$ χ^2 = 10.74, df = 2, p = 0.005).

3.9.3 Characteristics of study participants

Postal Questionnaire

The age, gender, posting, and rank category distributions of Postal Questionnaire respondents is shown in Table 3.5. There was a higher proportion of females in the exposed group compared to both comparison groups ($\chi^2_2 = 7.10, p = 0.029$), although this difference was small (2.4% versus 1% and 0.6%). Age ($\chi^2_{16} = 20.21, p = 0.21$) and rank ($\chi^2_4 = 1.04, p = 0.90$) were similar for the three exposure groups, but there were differences in posting category ($\chi^2_8 = 16.30, p = 0.038$). The main differences were in the proportion of respondents in each group with the earliest posting period (32%, 38% and 35% for Amberley comparison, Richmond comparison and exposed respectively).

Table 3.5 : Characteristics of eligible DSRS exposed and RAAF Amberley and Richmond comparison group Postal Questionnaire respondents

	Amberley N=489		Richmond n=592		Exposed n=635	
	n	%	n	%	n	%
Gender						
Female	3	0.6	6	1.0	15	2.4
Male	486	99	586	99	620	98
Age category						
≤ 24yrs	0	0	1	0.2	1	0.2
25-29yrs	20	4.1	20	3.4	31	4.9
30-34yrs	53	11	53	9.0	84	13
35-39yrs	93	19	91	15	106	17
40-44yrs	123	25	152	26	128	20
45-49yrs	99	20	137	23	129	20
50-54yrs	51	10	77	13	74	12
55-59yrs	34	7.0	42	7.1	50	7.9
60-64yrs	10	2.0	10	1.7	17	2.7
65-69yrs	6	1.2	7	1.2	11	1.7
70-74yrs	0	0	2	0.3	3	0.5
Posting category						
1975-1979	156	32	227	38	222	35
1980-1984	116	24	141	24	136	21
1985-1989	121	25	135	23	133	21
1990-1994	57	12	55	9.3	94	15
1995-1999	39	8.0	34	5.7	50	7.9
Rank category						
Civilian	0	0	0	0	39	6.1
Enlisted	305	62	366	62	368	58
Non-Comm. Officer	168	34	209	35	188	30
Officer	16	3.3	17	2.9	19	3.0
Unknown	0	0	0	0	21	3.3

Health examination

Table 3.6 shows the characteristics of individuals who participated in the health examination by group. The exposed group had a higher proportion of females (2.4%) than the comparison groups (0.3% and 1% for Amberley and Richmond respectively; $\chi^2_2 = 9.61, p = 0.008$), although these differences were small. Exposed individuals were more likely to be in the younger age categories than were the Amberley and Richmond comparisons ($\chi^2_{16} = 29.69, p = 0.02$). Posting category varied significantly across the three groups ($\chi^2_8 = 15.75, p = 0.046$ χ^2 = 15.75, df = 8, p = 0.046), with fewer exposed than comparison individuals with posting periods of 1980-1984 and 1985-1989. Apart from civilians, rank was similar for all three groups ($\chi^2_4 = 0.81, p = 0.94$).

Table 3.6 : Characteristics of DSRS exposed and RAAF Amberley and Richmond comparison groups health examination participants

	Amberley N=406		Richmond N=516		Exposed N=616	
	n	%	n	%	n	%
Gender						
Female	1	0.3	5	1.0	15	2.4
Male	405	100	511	99	601	98
Age category						
≤ 24yrs	0	0	0	0	1	0.2
25-29yrs	14	3.5	17	3.3	32	5.2
30-34yrs	41	10	48	9.3	81	13
35-39yrs	77	19	75	15	106	17
40-44yrs	107	26	134	26	123	20
45-49yrs	86	21	125	24	122	20
50-54yrs	42	10	68	13	68	11
55-59yrs	30	7.4	35	6.8	50	8.1
60-64yrs	6	1.5	7	1.4	18	2.9
65-69yrs	3	0.7	5	1.0	11	1.8
70-74yrs	0	0	2	0.4	3	0.5
Posting category						
1975-1979	134	33	191	37	216	35
1980-1984	102	25	130	25	132	21
1985-1989	96	24	121	23	128	21
1990-1994	46	11	47	9.1	94	15
1995-1999	28	6.9	27	5.2	46	7.5
Rank category						
Civilian	0	0	0	0	38	6.2
Enlisted	259	64	322	62	359	58
Non-Comm. Officer	135	33	179	35	179	29
Officer	12	3.0	15	2.9	18	2.9
Unknown	0	0	0	0	22	3.6

3.10 Discussion

General Health and Medical Study response rates for the Amberley and Richmond groups were disappointing (less than 50%), despite multiple attempts by mail, phone and in some cases email. After extensive follow-up, a large proportion of the selected sample remained uncontactable. This is not completely surprising given that some of the people we were attempting to contact had been posted up to 30 years previously and there is considerable movement of RAAF personnel with changes in postings, etc. A much smaller proportion of exposed individuals were not contactable. This is to be expected given that most of this group had been contacted recently as part of the Board of Inquiry or had contacted DVA to have their names included on the F-111 DSRS list.

Final response rates for any component of the study were 40% for Amberley, 48% for Richmond, and 77% for the exposed. Excluding all individuals who could not be contacted, response rates were 54%, 62% and 81% for Amberley, Richmond and the exposed group. However, it is likely that some of these non-contactable individuals were in fact passive refusers, and the true response rate is likely to lie somewhere between the rate that includes non-contactables and the rate that excludes them. Overall recruitment results for the current study (excluding deceased but including non-contactables) were slightly lower than those reported by the Monash University Gulf War Study[1] for exposed participants (77% versus 81%) and eligible comparisons (40% for Amberley and 48% for Richmond, versus 57%). Refusal rates for the current study were also higher for both the exposed group (18% versus 6.6% for the Gulf War Study) and comparisons (34% for Amberley and 30% for Richmond, versus 18%).

The three groups were initially well matched on gender, age group, posting category and rank category (apart from civilians). While this was to be expected given that the comparison groups were selected to match the exposed group in the distribution of these characteristics, the inclusion of the additional individuals reclassified as exposed (Round Three) did not distort this contrast.

The greatest concern for response bias was in the Richmond comparison group. Postal Questionnaire respondents and health exam participants differed substantially on age, posting category and rank category. Those included in the analyses were in the older age groups, earlier posting periods, and of higher ranks than non-respondents and non-participants. While there were some differences between Amberley respondents and non-

respondents, these were not consistent for the Postal Questionnaire and health examination. Health exam participants were in the older age groups and Postal Questionnaire respondents had higher ranks. Because of the very small number of females, gender differences are not clinically important. There were differences in rank for Postal Questionnaire respondents in the exposed group, but there was no other response bias for this group.

Although response bias was of concern, there were fewer differences between the three groups in the characteristics of Postal Questionnaire respondents and health examination participants. There was some indication that exposed individuals were from later posting periods than comparison individuals, but this difference was minor. There were also some differences in age between health exam participants. The differences in distribution of gender are so small as to be of no clinical significance. Age, posting, and rank categories will all be considered as potential confounders and included in all regression analyses.

3.11 References

1. Monash University. Australian Gulf War Veterans' Health Study. Vol I (Monash University, Melbourne, Vic, 2003).

4 Exposure

Chapter summary

The definition of “exposure” proved somewhat problematic for the General Health and Medical Study, and a final decision was made to base exposure on participation in a Deseal/Reseal (DSRS) program rather than by individual task or use of a particular chemical type. This chapter describes the Exposure Questionnaire used by the study team to collect data on program involvement and how this information was used to classify individuals as “exposed” or “non-exposed”. Final numbers are provided for the exposed group – overall figures as well as a break down by program type and by dose.

Chapter contents

4.1 Introduction

Whilst a toxicological assessment of the DSRS processes had been commissioned by the Board of Inquiry (BOI),[1] for the purposes of the General Health and Medical Study "exposure" was somewhat problematic to define. In occupational epidemiology an exposure is the "presence of a substance in the environment external to the worker"[2], with exposure levels assessed in reference to intensity and duration of a particular chemical. However, as seen from Table 4.1, a great variety of chemicals was used in the F-111 DSRS programs and they varied across tasks and across programs. Selecting which chemical(s) to focus upon would have been arbitrary since there was no *a priori* hypothesis (see Glossary) about which one(s) might have been responsible for the adverse events reported by DSRS personnel. In addition, focusing on one chemical would have overlooked the possibility that a particular combination of chemicals was responsible for adverse outcomes. At a round table Exposure Discussion Workshop (held on 12 March 2003) involving key decision-makers and advisers to the Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP)[*], it was decided that it was most appropriate for exposure to be defined at the level of a program. Hence all personnel who had participated in any of the official F-111 DSRS maintenance programs were included. This decision was based on the following reasons:

1) To ascertain participation in a program would be more accurate than trying to ascertain exposure to particular chemicals, especially given that the study was being done up to 30 years after the fact.

2) There would be less risk of misclassification for a definition based on program participation.

3) There would be less risk of type I error (i.e. concluding that there was an effect when there was no effect) because the number of analyses/comparisons would be reduced.

[*] Workshop attendees included: SAC representatives Dr Deborah Glass, Professor Michael Moore; DVA representative Mr David Goldrick; Defence Health Services Branch representatives GPCAPT James Ross, Dr Ian Gardner; RAAF representatives WO Peter Hind, FSGT George Cunningham; TUNRA representatives Associate Professor Catherine D'Este, Dr John Attia, Mr Richard Gibson, Ms Meredith Tavener, Mrs Maya Guest.

Formal F-111 DSRS programs included:

- Program 1, conducted between 1977 and 1982 (following training and preparatory works conducted at Sacramento, USA)
- Program 2, conducted between 1991 and 1993
- the Wing DSRS Program conducted from 1985 to 1992
- the Spray Seal Program which operated between 1996 and 1999.

“Pick and patch” (see Glossary) was an ongoing *ad hoc* activity which would have been very difficult to identify, so people who performed this task outside a formal DSRS program were not included in the current study. However, this does not necessarily imply that results from the Health Study will not be applicable to this group of individuals.

Table 4.1 : Summary of chemicals and processes for Deseal/Reseal activities

PROGRAM 1

Application	Product	Key Chemicals	Method of Application	Route of Exposure	Predicted Exposures[#]
Jet Fuel	JP-A-1 JP-8	Kerosene Aromatics (including benzene) 12-22% Additives <1%[3,4]	De-puddled from tank with rag	Respiratory Skin contact / absorption	Not assessed
Desealant	SR-51 SR-51A	Dimethyl acetamide Thiophenol Aromatic solvent	Mixing, handling, sprinkler	Respiratory Skin contact / absorption	Mixing & handling: low to medium Inside tank: high to excessive Hanger: low to medium
Alkaline / detergent wash	AIRTECH 23 ED-500	Glycol ethers (Thiophenol residues)	Mixing, handling, sprinkler Scrubbing solution with brush	Respiratory Skin contact / absorption	Liquid contact: low
Solvent	Mil-Spec	Naphtha Ethyl acetate Methyl Ethyl Ketone (MEK) Isopropanol	Spray bottle Wiped on with rag	Respiratory Skin contact / absorption	Inside tank: High to excessive liquid absorption

[#] As estimated by Connell, D. and Miller, G. (2001) Toxicological Assessment of Deseal/Reseal Chemicals - F-111 Fuel Tanks, Envirotest, Nathan, QLD.

Table 4.1 continued…

PROGRAM 1 continued…

Application	Product	Key Chemicals	Method of Application	Route of Exposure	Predicted Exposures
Metal surface protection	PR1560 DESOTO 823-707	Toluene Xylene Isopropanol	Wiped on with rag	Respiratory Skin contact / absorption	Not assessed
Metal surface protection	Alodine 1200S	Polyamide resin containing chromium trioxide	Wiped on with rag (anhydrous powder)	Respiratory Skin contact / absorption	Skin contact: low
2 Part epoxy barrier	XA3598 / (EC-3580 B/A)	Amine/amide resin Epoxy resin	Spray gun	Respiratory Skin contact / absorption	Skin contact: low to medium
Primer/adhesion promoter	PR-148	Naphtha Ethyl acetate MEK Isopropanol	Wiped on with rag	Respiratory Skin contact / absorption	Inside tank: high to excessive liquid absorption
Sealant	PRO-SEAL 899 PR-1750	Hydro-generated terphenyl Tricholoro-propane Manganese dioxide	Fillet or sealant injection gun	Direct contact with skin as a paste	Low to uncertain

Table 4.1 continued…

PROGRAM 2

Application	Product	Key Chemicals	Method of Application	Route of Exposure	Level of exposure
Jet Fuel	JP-4 JP-8	Kerosene Aromatics (including benzene) 12-22% Additives <1%[3,4]	De-puddled from tank with rag	Respiratory Skin contact / absorption	Not assessed
Solvent	ME767 / (T4460) MEK	Naphtha Ethyl acetate MEK Isopropanol	Spray bottle Wiped on with rag	Respiratory Skin contact / absorption	Inside tank: High to excessive liquid absorption
Metal surface protection	DESOTO 823-707	Toluene Xylene Isopropanol	Wiped on with rag	Respiratory Skin contact / absorption	Not assessed
Metal surface protection	Alodine 1200S	Polyamide resin containing chromic trioxide	Wiped on with rag (anhydrous powder)	Respiratory Skin contact / absorption	Skin contact: low
Barrier	EC-3580 B/A	Amine/amide resin Epoxy resin	Spray gun	Respiratory Skin contact / absorption	Contact: low to medium
Primer / adhesion promoter	PR-148	Naphtha Ethyl acetate MEK Isopropanol	Wiped on with rag	Respiratory Skin contact / absorption	Inside tank: high to excessive liquid absorption
Sealant	PR-1750	Hydro-generated terphenyl Tricholoro-propane Manganese dioxide	Fillet sealant injection gun	Direct contact with skin as a paste	Low to uncertain

Table 4.1 continued…

SPRAY SEAL PROGRAM

Application	Product	Key Chemicals	Method of Application	Route of Exposure	Level of exposure
Jet Fuel	JP-A-1 JP-8	Kerosene Aromatics (including benzene) 12-22% Additives <1%[3,4]	De-puddled from tank with rag	Respiratory Skin contact / absorption	Not assessed
Alkaline / detergent wash	ZI-400	Surfactants – anionic/nonionic	Mixing, handling, sprinkler Scrubbing solution with brush	Respiratory Skin contact / absorption	Mists / aerosols & liquid contact: high
Solvent	MEK	MEK	Spray bottle Wiped on with rag	Respiratory Skin contact / absorption	Inside tank: excessive levels (vapour and liquid) very high
Metal surface protection	Alodine 1200S	Chromium trioxide Potassium fluoborate Potassium ferricyanide (III)	Powder Wiped on with abrasive pad	Respiratory Skin contact / absorption	Skin contact: low
2 part Epoxy Barrier	EC-3580 B/A	Polyaminopolyamide	Fillet or sealant injection gun	Skin contact / absorption	Inside tank: high to excessive levels: inhalation and skin contact
Primer / adhesion promoter	666-2003-427 (MMS-425)	n-butyl acetate MEK Toluene Strontium chromate	Spray gun	Respiratory Skin contact / absorption	Inside tank: excessive levels, very high

Table 4.1 continued…

SPRAY SEAL PROGRAM continued…

Application	Product	Key Chemicals	Method of Application	Route of Exposure	Level of exposure
Sealant	PR-1750	Manganese dioxide Hydrogenated terphenyls Sodium polysulphide copolymer	Fillet gun Brush	Respiratory Skin contact / absorption	Potential skin contact: uncertain
Sealant	PR-2911	Toluene diisocyanate	Fillet gun Brush	Respiratory Skin contact / absorption	Inside tank: excessive levels for inhalation and skin contact
Sealant	PR-1826	Mercaptan polythioether polymer compound polymerctan	Fillet gun Brush	Respiratory Skin contact / absorption	Potential skin contact: uncertain

Table 4.1 continued…

WING DESEAL RESEAL PROGRAM

Application	Product	Key Chemicals	Method of Application	Route of Exposure	Level of exposure
Alkaline / detergent wash	ED-500 AIRTECH 23	Surfactants – anionic/nonionic	Ambient Level	Respiratory Skin contact / absorption	Not assessed
Solvent	T4460	MEK	Spray bottle Wiped on with rag	Respiratory Skin contact / absorption	Not assessed
Metal surface protection	Alodine 1200S	Chromium trioxide Potassium fluoborate Potassium ferricyanide (III)	Powder Wiped on with abrasive pad	Respiratory Skin contact / absorption	Not assessed
Barrier (epoxy adhesives)	EC-2216 EC-2580 B/A	Bisphenol A Kaolin	Sealant fillet gun	Respiratory Skin contact / absorption	Not assessed
Primer / adhesion promoter	EC-1945 B/A SS-4004 PR-148	Silicone polymer in solvents Acetone Alcohols Toluene Triethanol-amine	Wiped on with rag	Respiratory Skin contact / absorption	Low to high (personal breathing zone)
Sealant	PR-1750	Manganese dioxide Hydrogenated terphenyls Sodium polysulphide copolymer	Brush Fillet gun	Respiratory Skin contact / absorption	Not assessed
Sealant	Q4-2817	Fluorosilicone elastomeric sealant Silicanes Fluoro-silocane	Brush Fillet gun	Respiratory Skin contact / absorption	Inhalation: low to medium Skin contact: uncertain

4.2 Exposure Questionnaire

As described previously, individuals were initially assigned by the Department of Veterans' Affairs (DVA) to one of three exposure categories:

- Category 1 – directly involved with F-111 DSRS or had exposure to DSRS chemicals
- Category 2 – worked in close proximity to F-111 DSRS activities
- Category 3 – been at RAAF Base Amberley during the exposure period of interest.

In addition, individuals who registered with DVA after the DVA F-111 list lockdown on 12 September 2002 had not been allocated to an exposure category. The assignment to a category was done in a semi-structured manner by the DVA C&R Team interviewer according to their discretion. In order to clarify exposure status more rigorously, objectively and consistently, an Exposure Questionnaire (EQ) was developed; it comprised seven main sections (sections A-G) containing 67 items (see Appendix 2I). The EQ contained first-level screening items (Figure 4.1) which asked respondents to indicate if they had been involved in any of the F-111 DSRS programs (Program 1, Program 2, Wing Program, Spray Seal Program) or in the storage, mixing and/or disposal of DSRS-related chemicals. Where the respondent gave a positive response to any first-level item, they were then asked to provide further detail about their particular work activities, or tasks, through second-level questions (Figure 4.2). Finally, respondents were asked to identify the average time per week and the total number of weeks that they spent on these tasks.

A participant's handwritten comments were data-entered by the research team and used to inform the overall categorisation of that participant. Sections H and I gave respondents the opportunity to list additional duties and time periods, if their exposure did not "fit" into the previous sections.

Figure 4.1 : First level screening item from the Exposure Questionnaire

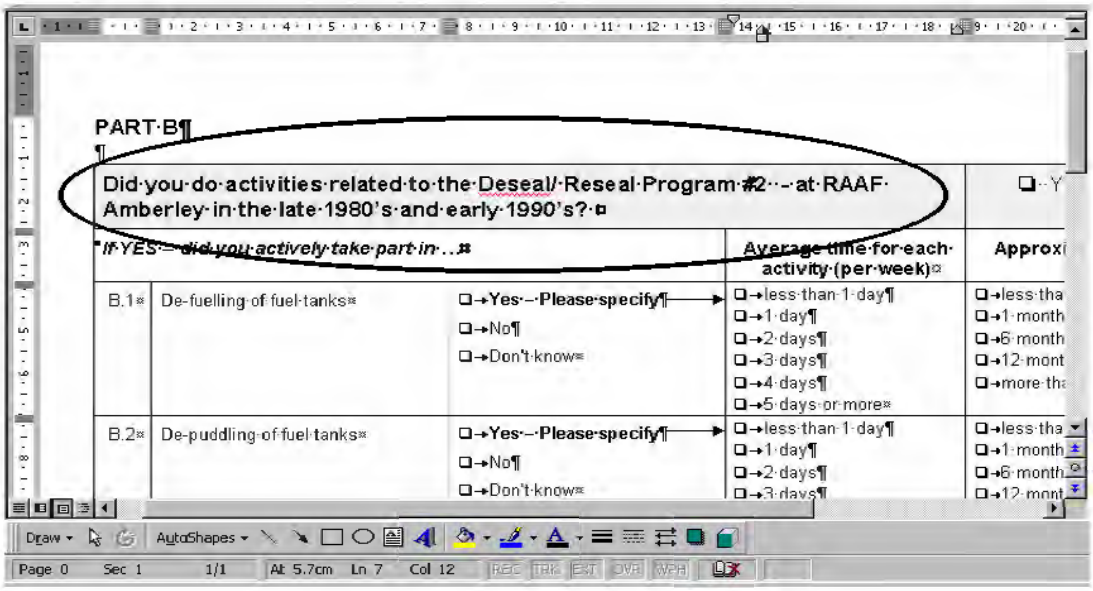
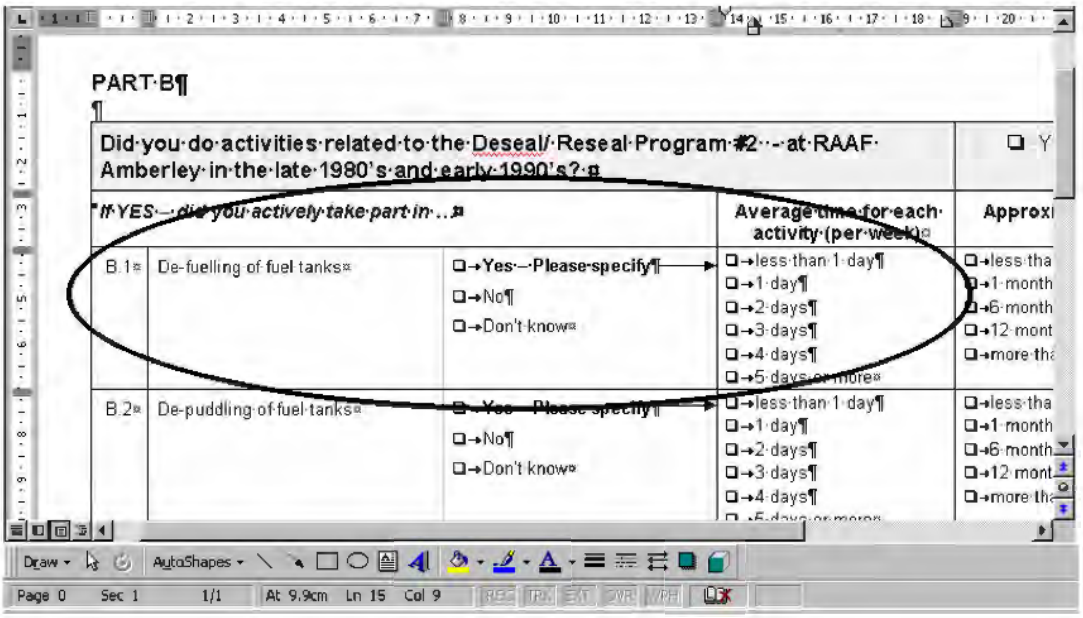


Figure 4.2 : Second level screening item from the Exposure Questionnaire



4.3 Classification

Based upon responses to the first- and second-level items in the EQ, each survey was reviewed by select members of the research team as described below.

[A] *Computer-coding methods.* Two people entered data from all questionnaires into an Access database. A participant was judged "exposed" if:

- a first-level screening item received a "yes" response
- it was indicated by responses to the second-level screening items that a respondent had undertaken F-111 DSRS activities
- it was indicated by responses to the "time of involvement" items that a respondent had undertaken F-111 DSRS activities.

[B] *Manual determination*. This was undertaken where there was no "yes" or "no" response to any first- or second-level screening item. In these cases:

- comments and/or attachments provided by the respondent were reviewed to see whether these indicated that they did take part in F-111 DSRS and related activities
- those who did return a questionnaire but did not provide sufficient information to allow classification, were contacted by the study team to facilitate completion of their EQ
- those who indicated that they were involved in mixing, storing or disposing of chemicals but who did not identify a particular program in which they undertook these activities were also coded manually. Where possible, assignment to a program was made on the basis of either the chemicals or the tasks mentioned. Three members of the research team were responsible for independently reviewing written comments and for deciding upon a respondent's final exposure status.

In cases where an exposed person could not be contacted or they had not completed an EQ, the exposure originally assigned by DVA from the F-111 DSRS database was retained. Further, given the level of agreement between returned EQ responses and those people DVA classified as "exposed", all those who were given a health examination (i.e. were initially classified as exposure category "1") but who did not return a questionnaire were also considered to be exposed for the purposes of this study.

Applying the standardised assessment of exposure (EQ results) to all those in the DVA F-111 database led to a number of individuals being reclassified (i.e. from the original DVA

category of 1, 2, 3 or no classification). A summary of the changes in exposure classification is shown in Table 4.2.

Table 4.2 : Exposure reclassification

Original Classification	Eligible and not known to have died	Returned Exposure Questionnaire		Classified as Exposed**	
	n	n	%	n	%
Category 1	705	491	*70*	454	*92*
Category 2	246	164	*67*	95	*58*
Category 3	242	119	*49*	48	*40*
Newly registered*	145	109	*75*	47	*43*
Total	1338	883	*100*	644	*100*

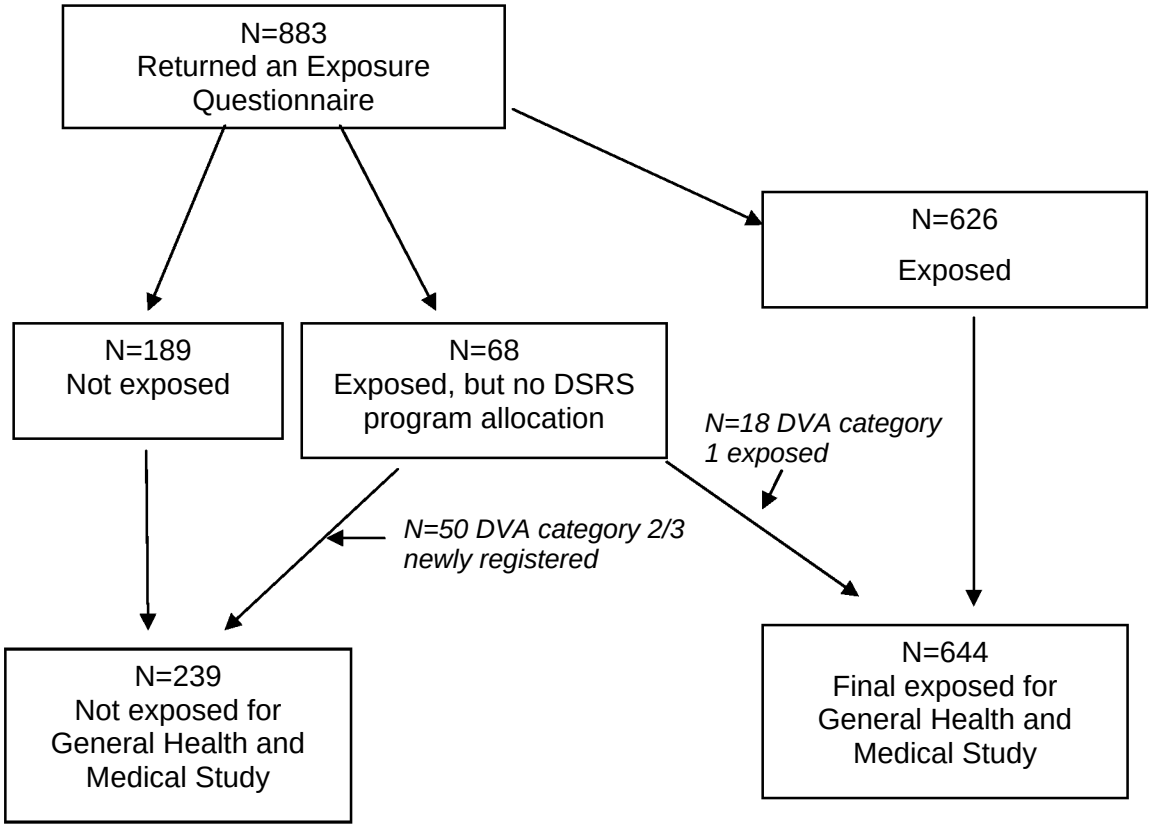
** 146 personnel registered after the September 2002 lockdown of the DVA F-111 list and were not categorised, one of whom died.*
******* *Includes 18 individuals from DVA category 1 who could not be allocated a program*

Overall, the EQ method of exposure classification appeared to be more sensitive and more inclusive than the procedures applied by DVA C&R personnel. A number of people originally classified by DVA as category 2 (worked in close proximity to DSRS) or category 3 (at RAAF Base Amberley but not exposed) were reclassified as “exposed” and became eligible to take part in the Health Study. Very few (37) originally classified as category 1 were reclassified as non-exposed.

4.4 Total numbers in exposed group

Figure 4.3 illustrates the final exposure classification process. A total of 883 participants returned an EQ (491 from DVA category 1, 164 from DVA category 2, 119 from DVA category 3, and 109 of the 146 individuals who registered between 12 September 2002 and 31 March 2003). Of these, 626 were classified as exposed and could be allocated to a DSRS program. A further 18 individuals with a DVA category of exposed were classified as exposed but could not be allocated to a DSRS program. These individuals retained a classification of exposed. An additional 50 individuals from DVA category 2, 3 and newly registered, indicated that they had been involved in DSRS activities, but they could not be allocated to a DSRS program. These individuals were classified as unexposed for the Health Study.

Figure 4.3 : Final exposure classification



There were 37 people who were eligible to participate in the health examination according to their original DVA classification of exposed (i.e. classified as category 1) but who were subsequently reclassified as “not exposed” following a review of their EQ. The following discussion refers to the 561 individuals who completed an EQ, who were classified as exposed, and who participated in a health examination.

4.5 Sub-group and dose characterisation

A consistent framework was adopted for all analyses. It was decided that the main analysis would include all exposed individuals (as defined above). Three sets of secondary analyses would be carried out: Program 1, Program 2, and a dose-response analysis. The rationale for these is outlined below.

4.5.1 Sub-group analyses

There was considerable overlap of involvement in the four programs and in the associated tasks of storage, disposal and mixing. This overlap can be observed by considering Table 4.3. The numbers in bold text represent the total number of individuals in a program. The non-shaded numbers represent the overlap between programs and activities. It can be seen that there were 293 people involved in Program 1 and 238 people involved in Program 2, with 64 people involved in both Programs 1 and 2. Therefore, 27% of those in Program 2 were also in Program 1, and 22% of those in Program 1 were also involved in Program 2.

Table 4.3 : Numbers of exposed personnel involved in more than one DSRS program and/or activity

Program	Program 1	Program 2	Wing Tank	Spray Seal	Storage	Disposal	Mixing
Program 1	**293**	.	.	.	.	.	.
Program 2	64	**238**	.	.	.	.	.
Wing Tank	74	153	**220**	.	.	.	.
Spray Seal	14	28	21	**80**	.	.	.
Storage	160	109	119	35	**264**	.	.
Disposal	122	34	44	9	122	**140**	.
Mixing	180	177	169	54	225	105	**368**

Additionally, there were 467 respondents involved in either Program 1 or Program 2, 279 respondents involved in Wing DSRS or Spray Seal Programs, and 444 involved in storage or disposal or mixing. Of those involved in the Wing DSRS or Spray Seal Programs, 72% were involved in either Program 1 or Program 2. In particular, 83% of those involved in the Wing DSRS and 41% of the 80 involved in the Spray Seal Program were involved in Program 1 or Program 2. Of those involved in the activities of storage, disposal or mixing, 79% were involved in either Program 1 or Program 2. Sub-group analyses were restricted to Program 1 and 2 because they had the greatest numbers of participants and hence had the greatest statistical power to detect an effect. This also captured most participants, in that only 14% of those in the exposed group who could be allocated to a program did not participate in either Program 1 or 2. There were too few participants in the Spray Seal Program to analyse separately.

4.5.2 Dose

The EQ required respondents to identify the average time per week and the approximate total number of weeks that they spent on each F-111 DSRS activity. It was determined at the Exposure Discussion Workshop that dose would be calculated using data collected on approximate total period of time involved only, because there was evidence that some respondents were confused about the “Average time for each activity (per week)” section. Within one program, the largest number of weeks reported for any task was taken as the time spent in that program. The frequency histograms in Figure 4.4 aggregate the maximum exposure time calculated for each individual for each program. For example, in Program 1 there was a total of 293 individuals, of whom 80 were involved with the program for more than 24 months, 60 were involved for 12 to 24 months, and nine provided no information about the time they were on the program. Similarly, there were only 80 people involved in the Spray Seal Program, with 31 being involved for more than 24 months and two people providing no information about exposure time.

Respondents could have been involved in more than one program, therefore total dose is an aggregate of dose from each program. The midpoint of each duration category was taken, and then the midpoints for each maximum time category (see Table 4.4) were summed across each program to create a total exposure time in months for each person. There were 531 individuals with a measure of exposure time.

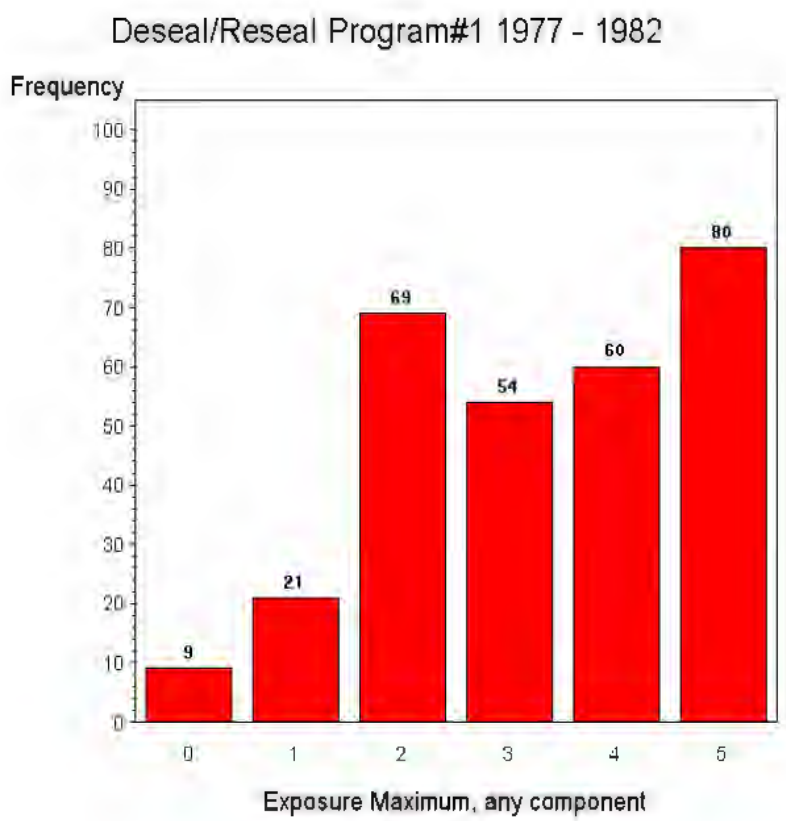
Table 4.4 : Exposure time midpoints

Approximate total period of time involved	Midpoint (months)
Less than 1 month	0.5
1 month to less than 6 months	3.5
6 months to less than 12 months	9
12 months to less than 24 months	18
More than 24 months*	30

** Based on inspection of the data, a midpoint of 30 months was considered reasonable for those who indicated more than 24 months of involvement.*

The resulting range of exposure time in months was from 0 to 120, with 30 people having no exposure recorded, 32 scoring the minimum exposure time of 0.5 months, and two individuals having the maximum possible time of 120 months (see Figure 4.5).

Figure 4.4 : Maximum exposure for any formal DSRS program component



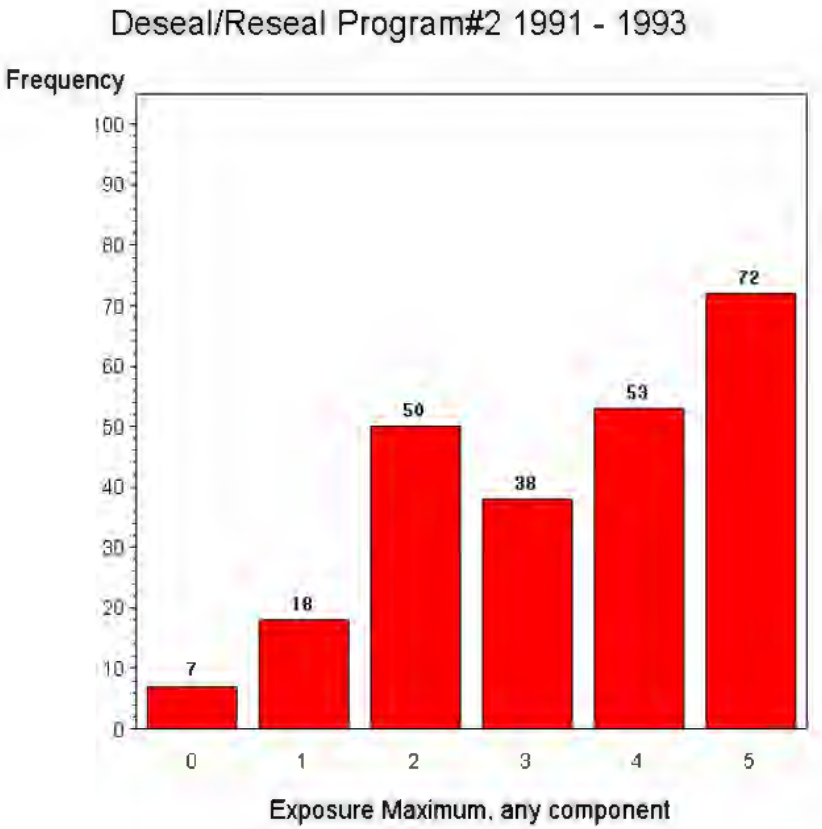
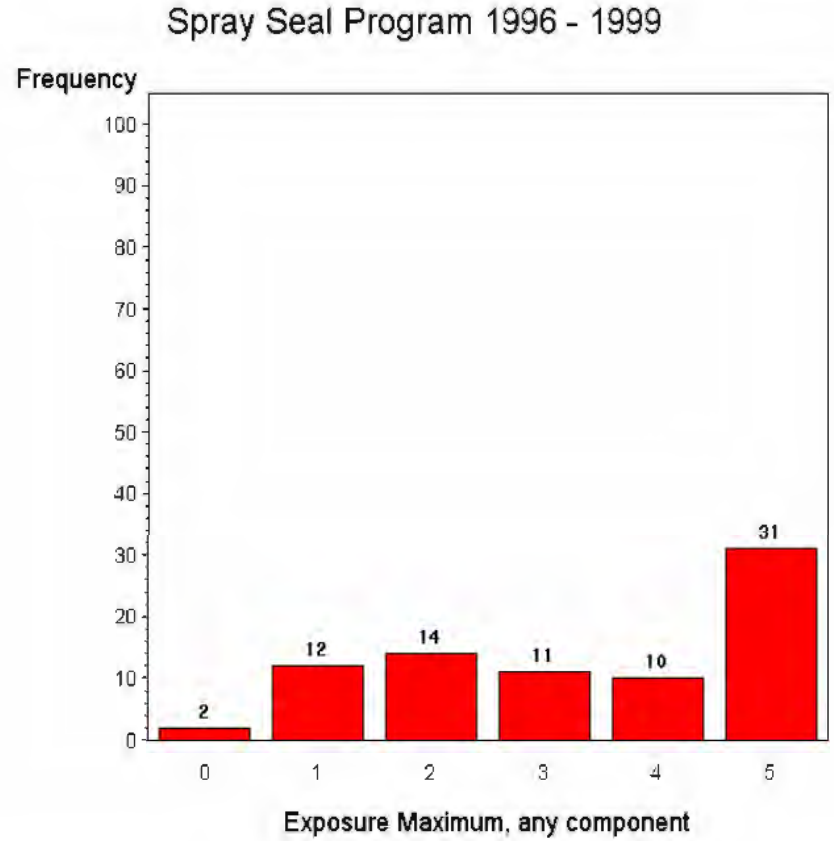
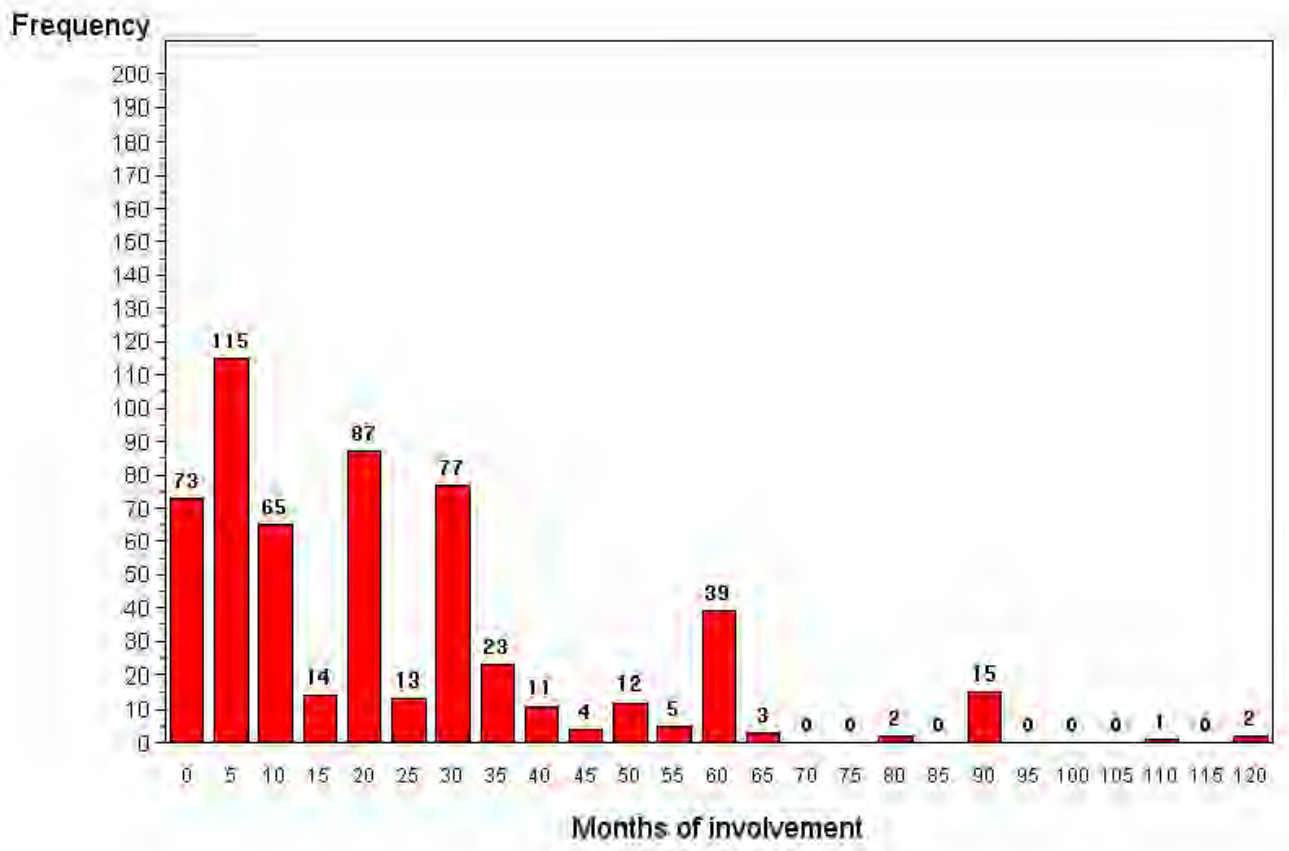


Figure 4.4 continued…



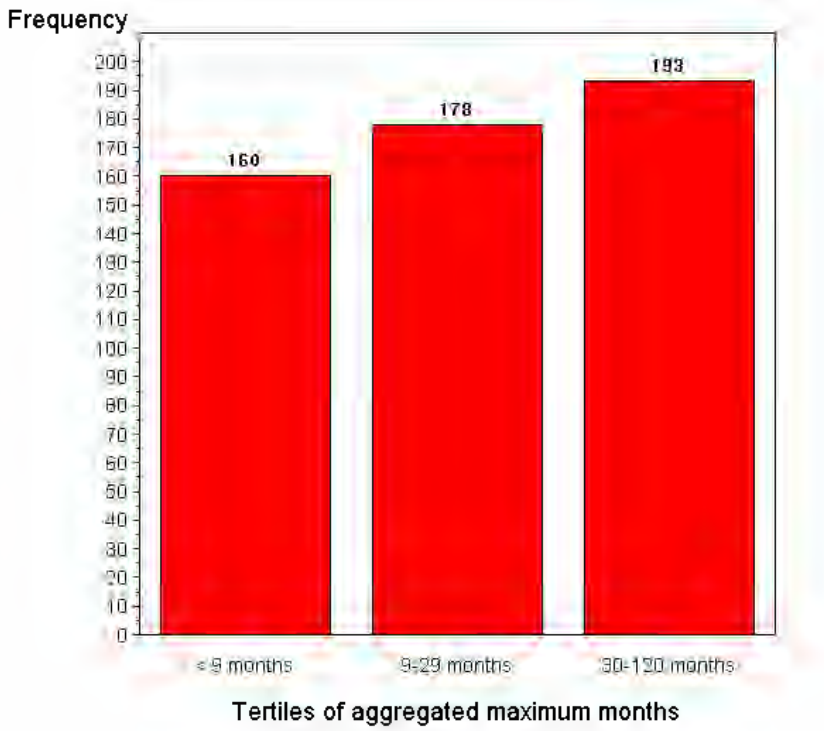
Code for exposure: 0= Not recorded, 1= Less than 1 month, 2= 1 month to less than 6 months, 3= 6 months to less than 12 months, 4= 12 months to less than 24 months, 5= Greater than 24 months.

Figure 4.5 : Maximum months of exposure summed across all programs



Excluding those with no exposure recorded, the median exposure time was 18 months and the mean was 21.7 months. The aggregate exposure time was then categorised into approximate tertiles, with 160 in the lowest exposure group, 178 in the next group, and 193 in the highest group (Figure 4.6). Points for division into tertiles were selected by accounting for the distribution of the data and considering clinically sensible end-points. Therefore, the current study will refer to the following three categories of exposure: mild (up to 9 months), moderate (10 to 29 months), and prolonged (30 months or more).

Figure 4.6 : Tertiles of maximum months of exposure



4.6 Discussion

This chapter describes the assignment of exposure status used in the rest of the report and provides the framework for analysis of exposure status used for the various Health Study outcomes. In summary, the definition of exposure using the EQ appears to be more consistent, objective, sensitive and inclusive than the more subjective scheme originally employed by DVA. The results reported in subsequent chapters will include the primary analysis using all exposed people from Programs 1 and 2, Wing DSRS and Spray Seal. Three secondary analyses will include those in Program 1, Program 2, and a dose-response comparison based on tertiles of time spent across all programs.

4.7 References

1. Connell, D. & Miller, G. Toxicological Assessment of Deseal/Reseal Chemicals - F1-11 Fuel Tanks. (Envirotest, Nathan, Queensland, 2001).
2. Checkoway, H., Pearce, N. & Crawford-Brown, D.J. *Research methods in occupational epidemiology*, (Oxford University Press, New York, 1989).
3. Ritchie, G. et al. Biological and health effects of exposure to kerosene-based jet fuels and performance additives. *J Toxicol Environ Health B Crit Rev* **6**, 357-451 (2003).
4. Vere, R.A. Aviation fuels. in *Modern Petroleum Technology, Part 2* (ed. Hobson, G.D.) 723-771 (John Wiley & Sons, Chichester, 1984).

5 Measures

Chapter summary

This chapter provides a summary of the type of health information collected from participants in the General Health and Medical Study. Self-reported data were collected via a Postal Questionnaire covering topics such as general health and well-being, medications, alcohol and smoking behaviours, memory, sexual health, and occupational history. In addition, a comprehensive physical and psychological examination was also conducted by Health Services Australia including tests for cardiovascular and respiratory health, dermatological and breast abnormalities, vision, smell, hearing, vibration sensation and tremor, blood pressure, mood, memory, and learning.

Chapter contents

5.1 Introduction

The General Health and Medical Study (the Health Study) in the Study of Health Outcomes in Aircraft Maintenance Personnel project (SHOAMP) involved the collection of self-reported health and occupation details via a Postal Questionnaire, as well as objective physical measures undertaken in a comprehensive health assessment conducted by health professionals at Health Services Australia (HSA) centres throughout selected regions of Australia. Appointments were arranged by the Department of Veterans' Affairs (DVA) Contact and Recruitment (C&R) Team. Every HSA team consisted of administrative and data entry personnel, medical practitioner, psychologist and nurse, each of whom were provided with training and instructional manuals which outlined all the procedures to be followed for SHOAMP (see Chapter 7 for more detailed descriptions).

5.2 Postal Questionnaire

The Postal Questionnaire was an instrument for self-completion developed specifically for SHOAMP participants. It was posted out to each potential participant with their original invitation package for the Health Study, in addition to general information about the study, an Exposure Questionnaire (for exposed personnel) and a Reproductive Questionnaire (for female partners of male participants). Each questionnaire contained an introductory page of free-call 1800 phone numbers to contact for more information or assistance, together with a two-page consent form which summarised each aspect of the study for which consent would be required. Participants were asked to complete the Postal Questionnaire and consent form in their own time and bring the completed forms with them to their health examination with Health Services Australia. A reply-paid envelope was also included in the package so that participants who chose not to have a health examination could post their completed questionnaire to the study team. To help participants complete the Postal Questionnaire, assistance was provided over the telephone by the study Project Manager or Investigators.

The Postal Questionnaire (Appendix 2K) was divided into 12 main areas of investigation:

1) general demographic items to check date of birth, gender, marital status, and the highest qualification reached

2) a list of various symptoms that participants might experience (informed by anecdotal reports of poor health to the DVA F-111 Interim Health Care Scheme and by studies of Gulf War participants)

3) a list of possible conditions for which participants might have received a positive diagnosis

4) items regarding hospitalisation, family history of selected psychological conditions, and malignancies

5) medications currently being taken by the participant (where both prescription and over-the-counter medicines could be recorded), reclassified according to generic name using the MIMS online facility (http://www.mims.hcn.net.au, last accessed 11 February 2004), then coded further according to the World Health Organization (WHO) Anatomical Therapeutic Chemical (ATC) classification system (http://www.whocc.no/atcddd, last accessed 11 February, 2004)

6) lifetime alcohol intake

7) lifetime smoking behaviour

8) memory

9) quality of life (using the SF-36 quality of life instrument)

10) occupational history (where questions were asked regarding civilian and defence employment)

11) a section specific to some of the tasks carried out during aircraft maintenance

12) sexual function (for male participants) or reproductive history (for female participants and female partners of male participants).

5.2.1 Selection of Postal Questionnaire measures

In addition to input from The University of Newcastle Research Associates (TUNRA) study team members and project consultants, key information and expert sources used to inform the content of the Postal Questionnaire included:

- extensive literature review of adverse health outcomes potentially associated with DSRS, undertaken during Phase I of SHOAMP[1]
- report on a series of qualitative interviews with DSRS workers as part of Phase I of SHOAMP[2]
- Scientific Advisory Committee members

- Monash University Gulf War Study
- Gulf War Illness Research Unit at Kings College Hospital, London[3,4]
- Hopkins Symptom Checklist[5]
- F-111 DRS Interim Health Care Scheme
- project stakeholders.

Table 5.1 provides details of each health topic included in the Postal Questionnaire as well as the rationale for its inclusion. More details are provided on each measure in the relevant outcome chapter.

Table 5.1 : Rationale for Postal Questionnaire items

Area of Investigation – Postal Questionnaire	Rationale for inclusion
Section 1: Socio-Demographic **Items 1.1 – 1.5** **Gender** **Date of birth** **Country of birth** **Marital status** **Education level**	This section provides data on potential confounders, and values of some outcomes are likely to be associated with age, gender. If answered by non-consenting individuals, can provide basis for comparison between participants and non-participants.
Section 2: Personal health **Items 2.1 – 2.80** **Recent health symptoms**	Included due to the range of poor health symptoms being reported to the F-111 Interim Health Care Scheme by DSRS workers, and clusters of symptoms reported in the solvent exposure literature. Informed by the 63-item Monash Gulf War Study checklist, which in turn was based on items used by the Gulf War Illness Research Unit at Kings College Hospital (London) and the Hopkins Symptom Checklist. Some specific items not included for SHOAMP were “intolerance to alcohol”, “skin infections”, “shaking”, tingling or burning sensation in hands or feet”, “loss of sensation in hands or feet”, “lump in throat”, and “burning sensation in the sex organs”. Additional items included were “bleeding during bowel movements” and “irregular” added to the item on heart beat. Items for “dizziness”, “fainting”, “blackouts” were kept as individual items rather than combined as one item. Final number of items = 80.

Table 5.1 continued…

Area of Investigation	Rationale for inclusion
Section 2: Personal health cont… **Items 2.81 – 2.127 (male version)** **Items 2.81 – 2.129 (female version)*** **Diagnosed or treated conditions**	This section serves to collect data on existing conditions. Allows comparison of health between groups when participants consent to questionnaire only. Items removed for SHOAMP were “or condition” from heart disease, “diarrhoea”, “constipation”, “incontinence or difficulty passing urine”, “malaria”, “any significant infections”, “eye or vision problems”, “sick building syndrome”, “low fertility”, “other skin cancer”, “any other kind of cancer”, “disease of the hair or scalp”, and “drug abuse”. “A thyroid problem” was changed to “hyper-thyroidism” or “hypo-thyroidism”. “Blood disorder” was changed to “leukemia” and “lymphoma”. “Anxiety” and “depression” were split into two separate items. *Questions for women only regarding menstrual problems were shifted to the Female Version only (items 2.128 and 2.129). Two additional items were included to specify types of diagnosed cancers and any type of psychiatric or psychological condition (items 2.128 and 2.129 for the male version and items 2.130 and 2.131 for the female version).
Item 2.130 (male version) **Item 2.132 (female version)** **Hospitalisation**	Single hospitalisation item can provide an indication of health need and service utilisation between exposed and comparison groups.
Item 2.131 (male version) **Item 2.133 (female version)** **List of current medications**	This item provides an indication of health need and can provide indication of certain conditions (i.e. depression). Medications were reduced to name of medication only, due to questions regarding the quality of data that would be received if more information were asked. Only those medications needing a clinician prescription were included for analysis.
Section 3: Family medical history **Items 3.1 – 3.3** **Parents** **Siblings**	These items provide essential information on familial risk; these are potential confounders for some study outcomes. For SHOAMP, Family Medical History was simplified and contained more generic references: asking if immediate family had suffered from “depression”, “dementia” or “Alzheimer’s disease”. Item 3.2 asked if immediate family member has had cancer. Item 3.3 asked which family member and type of cancer.

Table 5.1 continued…

Area of Investigation	Rationale for inclusion
Section 4: Behavioural factors **Items 4.1 – 4.9** **Alcohol intake habits** **Items 4.10 – 4.20** **Lifetime alcohol consumption diary** **Items 4.21 – 5.24** **Smoking habits** **Item 4.25** **Lifetime smoking diary**	Smoking and alcohol intake are potential confounders for many outcomes. In particular, collection of lifetime habits is recommended as part of the overall assessment of DNA damage. Smoking in particular is also a sensitive issue in terms of poor health outcomes including cancer, and an estimation of "lifetime" consumption is desirable. Format for cumulative history of alcohol and smoking intake was based on Study of Childhood Cancer being conducted by researchers from the Hunter Medical Research Institute. This study in turn developed forms based on a Canadian study of lifetime drinking.
Section 5: Memory **Items 5.1 – 5.6** **Memory Complaint Questionnaire (MAC-Q)**	Subjective memory assessment was included due to evidence of adverse affects on neuropsychological functioning from solvent exposure. It also reflects complaints of adverse health symptoms by F-111 DSRS personnel, especially issues with poor memory. It was initially proposed that subjective memory be assessed by the 43-item Subjective Memory Questionnaire, however it was agreed that the much shorter 6-item Memory Complaint Questionnaire (MAC-Q)[6] scale would be included in the Postal Questionnaire as a self-completed test of memory.
Section 6: Quality of life **Items 6.1 – 6.11** **Short Form 36-item quality of life survey (SF-36)**	A measure of general well-being was included in the Postal Questionnaire as DSRS exposure may have had a negative impact on quality of life; also anecdotal evidence of social isolation due to DSRS involvement (ie: unpleasantness of tasks, resulting smell etc). The SF-36 is a valid and reliable instrument which is internationally recognised and has been used with a variety of settings and populations.

Table 5.1 continued…

Area of Investigation	Rationale for inclusion
Section 7A-7C: Occupational Exposures **Items 7.1 – 7.18** **Job types held, length in job, potential exposure to chemicals** **Item 7.19** **Aircraft maintenance checklist**	Information on civilian occupation was important as exposure to solvents and other chemicals was a potential confounder. This section was informed by previous studies which asked participants details of different types of exposures from different jobs held before and after Australian Defence Forces experiences. This section was based on respondent descriptions of "job title and description of work duties" and "industry type". Participants were not expected to name their exact employers, nor specific types of exposures. The maintenance checklist was included to provide information on all types of aircraft maintenance that participants may have performed (in and outside of ADF experiences).
Section 8: Sexual Function (male version only) **Items 8.1 – 8.15** **International Index of Erectile Function**	The issue of sexual function was included in SHOAMP due to several cases of self-reported sexual dysfunction by DSRS personnel. These items were in the Male Version only of the PQ. Physical sperm sampling and erectile function testing were deemed inappropriate for SHOAMP, and a decision was made to include questionnaire items only for each participant to self-complete. The International Index of Erectile Function[7] was included in the PQ, a 15-item self-complete questionnaire for males only.
Section 8: Reproductive Health (female version only) **Items 8.1 – 8.15** **Pregnancy, Miscarriages, Fertility** **Alcohol, smoking, caffeine behaviours**	This section (only in the female version) was included in the PQ to investigate concerns regarding effects of exposure on attempted or actual pregnancies during time of exposure, and to collect some data on health habits during each pregnancy. No/yes categories of response were included for any alcohol, smoking or caffeine behaviours during any pregnancy.

Sections 9 to 11 provided contact details for each participant, contact details of their preferred medical practitioner and the opportunity to provide feedback to the study team on SHOAMP processes and documentation respectively. There was also a separate confidential questionnaire included in the mail-out to male participants, to pass onto a female partner for completion. This questionnaire contained the same content as Section 8 in the Female Version of the PQ regarding reproductive health.

5.3 Health examination

Appointments for each health examination were made by personnel from the DVA C&R Team using the Electronic Management of Medical Assessments (ELMA) diary system link between DVA and HSA. Appointments were scheduled for the morning or afternoon; participants were told to expect their assessment to take approximately four hours. Prior to each health examination a reminder package was sent to participants which included directions to the nominated HSA office, confirmation of the time and date of their appointment, instructions about taking medications, not smoking and/or drinking alcohol prior to their examination and the need not to fast. DVA personnel telephoned each participant 24-48 hours prior to their examination as a further reminder. If a participant did not arrive for their health examination appointment, HSA personnel were instructed to call the DVA 1800 number or to send an email with the person's ID number. A follow-up call was then made, and a second appointment was scheduled wherever possible.

5.3.1 Summary of health examination procedures

Aside from the psychological evaluation, health areas to be assessed were allocated evenly between HSA nursing and medical staff. The aim was to have participants spend a similar amount of test time with each staff member so as to reduce waiting time between tests. Participants were asked to bring along their completed Postal Questionnaire; whilst they were having their health examination, this was data-entered, and before they left HSA any areas of missing data were checked with them. It was anticipated that the complete health assessment would take approximately three to four hours. All health examination data were recorded directly into a participant Health Examination Booklet (see Appendix 2N). Queensland Medical Laboratories (QML) undertook analyses of all blood samples; blood samples collected in health examinations conducted outside Brisbane were transported by World Courier.

5.3.2 Blood collection

HSA nursing staff were responsible for the collection of blood from each SHOAMP participant. Where a participant did not give consent for blood collection, a note was made in the Health Examination Booklet by nursing staff and cross-checked against the person's consent form. A single pathology service, QML, was contracted by DVA to process all results

for the study. QML provided each HSA office with blood collection kits labelled especially for SHOAMP participants. Each kit included:

- bar-coded labels for request form and collection tubes
- pathology request form
- 8.5 ml SST tube for biochemistry
- 4.5 ml EDTA tube for full blood count
- “soft-draw” collection tube for EBV infection
- blood slide and transport container.

Blood collection from the Ipswich and Brisbane offices occurred twice daily – morning and early afternoon – so that blood was collected and spun within a two-hour period. Other HSA offices spun blood before it was dispatched once only in the afternoon. World Courier was used to transport the blood from interstate offices to QML for processing. All results were sent to the relevant HSA office to be attached to the participant’s Health Examination Booklet.

5.3.3 Selection of health examination measures

The inclusion of specific health topics and/or tests as part of the General Health and Medical Study was based on findings from a review of the scientific and occupational health literature, in addition to a review of the symptoms being reported to the F-111 DSRS Interim Health Care Scheme by Air Force personnel involved in the DSRS programs. A brief rationale for each health area of examination is presented in Table 5.2, with each measure detailed further in the chapter to which it pertains. All health examinations were conducted by professionals at HSA offices across Australia. All staff members were “blinded” to the exposure status of each participant (i.e. whether they were in the exposed or comparison groups). They were asked to record whether or not the exposure status became evident at any point during the health examination, and if so, how (i.e. the participant described their DSRS work experiences).

Table 5.2 : Rationale for health examination items

Area Of Investigation – Health Examination	Rationale for Inclusion
HSA NURSING PERSONNEL	
Biomarkers, Full Blood Count, Liver Function Test, Electrolyte + Urea, Creatinine, Glucose, Cholesterol, Ca And K+, C-Reactive Protein, Apoliproprotein E, Homocysteine	Full blood count as part of comprehensive health examination. C-reactive protein can be a non-specific marker of inflammation. Liver function test for hepatic toxicity. Elevated calcium may be an indicator of cancer. Increased plasma Homocysteine level and Apoliproprotein E identified as risk factor for Alzheimer's Disease.
Height / Weight **Urinalysis**	Part of general comprehensive health examination. Height and weight for calculation of Body Mass Index (which is a potential confounder for some outcomes), and for use as part of lung function. Urinalysis to check for protein, blood and/or glucose in urine as an indicator of renal disease.
Visual Acuity	Included as part of overall health examination to confirm any vision deficits and used as a potential confounder in colour vision outcomes. Tested by Snellen chart at 6 metres distance.
Colour Vision **Ishihara Colour Plate Test** **Farnsworth D15 (saturated test)** **L'Anthony D15 (desaturated test)**	Three tests of colour vision were included for SHOAMP to identify general colour vision deficits, and also the loss of blue/yellow colour vision, previously reported in the literature as being affected by solvent exposure.[8]
General Health Questionnaire 12-item	To screen for short-term changes in mental health: depression, anxiety, social dysfunction and somatic symptoms.
Hearing	Hearing was included as part of overall comprehensive health examination to confirm any hearing deficits. Both bone and air conduction testing were included.
Respiratory	Lung function testing by spirometry, including pre and post bronchodilator measurement.
Foecal Occult Blood Test (FOBT)	SHOAMP recognised the reports by F-111 DSRS personnel of colonic spasms, polyps and colon cancer. The inclusion of a take-home FOBT kit also coincides with general preventive recommendations.[9]

Table 5.2 continued…

Area Of Investigation	Rationale for Inclusion
HSA MEDICAL PERSONNEL	
Mini Mental State Examination	As a measure of general cognitive impairment.
Pulse Rate **Chest Sounds**	Assessment of pulse rate and chest sounds made up an overall health examination for comparison personnel, and was also in response to adverse respiratory symptoms reported by F-111 DSRS personnel. Pulse (at the radial site) and blood pressure were measured in three positions: seated position, lying and standing; with the lying and standing measures used to assess postural hypotension. Addition of auscultation of chest sounds added to the overall expectation by participants of what constitutes a medical examination, and represented standard testing by HSA. For abnormal sounds, doctors recorded the presence of decreased breath sounds, bronchial breath sounds, wheezes, rubs or crackles as appropriate.
Blood Pressure	Blood pressure measurement formed part of overall health examination. Blood pressure was measured twice while seated, once while lying and then standing. Doctors recorded direct ausculatory measurement of systolic and diastolic pressure and postural drop.
Kessler 10-item scale	Included as a general measure of anxiety and depressive symptoms, it was administered between the first and second seated blood pressure readings.
Vibration Sensation (Neuropathy)	The measurement of vibration sensation was an important issue as organic solvents are capable of causing or contributing to peripheral neuropathies. Significantly, different vibratory threshold perceptions have been reported in long-term solvent exposed painters.[8,10] Measurement was by Biothesiometry, with three lower limb points tested (dorsal surface of the big toe, medial malleolus, medial side of the knee) and four upper limb points tested (middle finger distal interphalangeal joint, middle finger metacarpophalangeal joint, radial styloid, olecranon).
Balance	The Functional Reach Test was used as an indication of balance, as there had been some reports of “balance disorders” from DSRS personnel, and some literature reported dose-related effects on balance from solvent exposure.

Table 5.2 continued…

Area Of Investigation	Rationale for Inclusion
HSA MEDICAL PERSONNEL continued…	
Skin Analysis	DSRS personnel had reported a number of skin problems (i.e. rash, itchiness, dryness, peeling), which warranted the inclusion of a skin examination. This also formed part of a comprehensive health examination. Doctors were asked to report lipoma, psoriasis, dermatitis and “other” skin lesions.
Breast Examination **Lymphadenopathy**	One DSRS worker had been previously diagnosed with breast carcinoma and a second had previously reported breast enlargement. Clinical breast examination for all participants (male and female) was supported, to reduce the potential for diagnostic suspicion bias. A check was made of the lymph nodes in the neck (cervical) and supraclavicular fossae. Abnormalities were noted as present or absent.
Tremor	The literature reports tremor following exposure to solvents. Two tests were conducted for SHOAMP: the Groove Type Steadiness Tester (where a hand-held stylus was moved horizontally from left to right along the narrowing groove without touching the sides of the groove) and the Nine-hole Steadiness Tester (which used a metal plate with nine holes of diminishing size where the stylus had to be inserted without touching the sides).[11]
Olfactory **Sniffin’ Sticks Test**	There had been reports by DSRS personnel of a loss of sense of smell and malodorous (unpleasant) smell.

Table 5.2 continued…

Area Of Investigation	Rationale for Inclusion
HSA PSYCHOLOGICAL PERSONNEL	
National Adult Reading Test	Measure of pre-morbid intellect. The inclusion rationale for all neuropsychological assessments was that the literature provides evidence of adverse affects on neuropsychological functioning from solvent exposure. Also, it recognises complaints of adverse health symptoms by F-111 DSRS personnel, especially with poor memory.
Rey 15-item Test	To screen for the validity of neuropsychological test results.
WAIS III Similarities Test **Controlled Oral Word Association Test** **Trail Making Test B**	Measures of Executive Functioning which includes the ability to organise thoughts and work as part of day-to-day living, impulse control, resistance to distraction, self-awareness across time; questioning and reading comprehension; and self-regulation of emotion and motivation.
WAIS III Digit Symbol Coding **Trail Making Test A** **Purdue Pegboard**	Measures of Psychomotor Speed – which is the amount of time it takes a person to process a signal, prepare a response and execute that response.
Digit Span Forwards **Digit Span Backwards**	Measures of Attention / Working Memory – the ability to integrate and manipulate new information.
WAIS III Incidental Learning **WMS Visual Reproduction** **Auditory Verbal Learning Test**	Measures of New Learning and Memory – the ability to absorb, store and recall new information after experiencing a delay.
Block Design Test	Measure of Visuospatial abilities – spatial problem-solving and manipulative abilities.
CIDI Depression module	Measure of Depression
CIDI Anxiety module	Measure of Anxiety
National Mental Health module	Measure of Neurasthenia

5.4 References

1. TUNRA. The Study of Health Outcomes in Aircraft Maintenance Personnel. Phase I. Literature Review. (The University of Newcastle Research Associates Ltd, Newcastle, 2003).
2. TUNRA. The Study of Health Outcomes in Aircraft Maintenance Personnel. Phase I. Qualitative Interviews. (The University of Newcastle Research Associates Ltd, Newcastle, 2003).
3. Ismail, K. et al. Is there a Gulf War syndrome? *The Lancet* **353**, 179-182 (1999).
4. Unwin, C. et al. Health of UK servicemen who served in Persian Gulf War. *The Lancet* **353**, 169-178 (1999).
5. Derogatis, L., Lipman, R., Rickels, K., Uhlenhuth, E. & Covi, L. The Hopkins Symptom Checklist: A self report symptom inventory. *Behavioural Science* **19**, 1-15 (1974).
6. Crook, T.H., Feher, E.P. & Larabee, G.J. Assessment of memory complaint in age-associated memory impairment: The MAC-Q. *International Psychogeriatric* **4**, 165-175 (1992).
7. Rosen, R.C. et al. The International Index of Erectile Function: A multidimensional scale for assessment of erectile dysfunction. *Urology* **49**, 822-830 (1997).
8. Dick, F., Semple, S., Chen, R. & Seaton, A. Neurological deficits in solvent-exposed painters: a syndrome including impaired colour vision, cognitive defects, tremor and loss of vibration sensation. *Quarterly Journal of Medicine* **93**, 655-661 (2000).
9. *Guide to Clinical Preventive Services. Report of the U.S. Preventive Services Task Force*, (Williams and Wilkins, Baltimore, 1996).
10. Hageman, G. et al. Parkinsonism, pyramidal signs, polyneuropathy, and cognitive decline after long-term occupational solvent exposure. *Journal of Neurology* **246**, 198-206 (1999).
11. Louis, E.D. et al. Validation of a portable instrument for assessing tremor severity in epidemiologic field studies. *Movement Disorders* **15**, 95-102 (2000).

6 Analysis

Chapter summary

This chapter describes the approach taken for all General Health and Medical Study analyses. Exploratory data analyses were conducted for each study outcome: continuous variables with a non-normal distribution were transformed or dichotomised, and categorical variables with more than two categories were also dichotomised. For health examination outcomes, variability across Health Services Australia (HSA) centres was examined using a test of heterogeneity. Analysis of variance was used for continuous variables and the Breslow-Day test for dichotomous variables. Results for each outcome are presented as two tables: (a) a descriptive table which provides summary information by group, and (b) a summary regression analysis table which includes (i) results for the primary analysis of the exposed group versus the two comparison groups, and (ii) results for the secondary analyses, including Program 1 and Program 2 sub-groups, and the dose-response relationship.

Chapter contents

6.1 Approach to analysis

This chapter outlines the general approach undertaken for all analyses. Any deviations from this protocol or any additional analyses are reported in the relevant chapter(s). In general the following steps were taken:

- exploratory data analysis
- checking for variability across HSA centres
- descriptive analysis
- primary analysis
- secondary analyses.

6.2 Data management

Data were entered into a Microsoft Access database (there were several databases for the different study components). All data were double-entered, and a series of checks was undertaken to compare all numeric variables from the two data sets. Discrepancies were then checked against the original documents and amended as appropriate. Logic checks were undertaken in a similar manner. Details of data checks are provided in the chapter on quality assurance (see Chapter 7).

Statistical analyses were undertaken using the STATA and SAS statistical software programs. With the exception of data tables for female reproductive health (Chapter 14), all tables of analysis results were produced directly from programming of SAS output, without the need for any data to be re-entered into individual chapter tables. This reduced the possibility for transcription errors. Further quality assurance procedures and checks are detailed in Chapter 7.

6.3 Exploratory data analysis

Exploratory data analyses were undertaken for each outcome. For categorical variables, frequency distributions were obtained and checked for missing and out-of-range values. For continuous variables, summary statistics and graphs were produced and assessed for missing values, potential outliers, and normality of distributions. Any unusual values were

checked against original data collection forms. Values which could not be verified were set to "missing".

For continuous variables with a distribution that was not (approximately) normal, various transformations were tried in an attempt to normalise the data. The type of transformation used depended on the shape of the original distribution, but more particularly, was chosen to ensure normalisation and stabilisation of the residuals of the final analysis of variance model. If a normal distribution could not be obtained, the outcome was then dichotomised, with the cut-point based on either clinical protocol, on the 10^{th} percentile of the distribution of population norms, or the 10^{th} percentile of the Richmond comparison group (as this was considered to be the most appropriate relevant comparison group in the absence of population norms). This categorisation generally occurred for outcomes which had a substantial floor or ceiling effect (i.e. a large proportion of the participants had the highest or lowest possible value for the variable).

For categorical variables with more than two categories (for example, variables which provided level of severity for an outcome such as none, mild, moderate or severe), categories were collapsed to form dichotomous variables for analyses. All severity categories were combined to form a variable which indicated the presence or absence of the condition of interest. Results of this exploratory analysis are reported in the text of each chapter. Where data were dichotomised and one of the categories comprised less than 5% of the data, no further analysis was undertaken.

6.4 Variability across HSA centres

Due to the fact that participants for the current study were recruited from across Australia, it was not possible to perform the health examinations at one centralised location with one team of clinicians. Multiple centres and multiple teams were used. This was an advantage in terms of efficient use of time and resources, but a disadvantage in that there was potential for inter-clinician variability in diagnosis, as well as variation due to differences in type, calibration, and use of equipment. The study team sought to minimise this variability by providing extensive training to all clinicians prior to the study's commencement and by establishing other quality assurance procedures (see Chapter 7). However, the possibility remained that for outcomes which were highly subjective, such as gynaecomastia or lipoma assessed during a physical examination, there may have been continuing variability. In

addition, there was variability in the relative numbers of participants from the three exposure groups attending each of the HSA centres. For example, the proportion of participants attending each HSA centre who were in the exposed group varied (see Chapter 8).

For this reason, two checks were put into place at the second stage of the analysis:

1) It was possible that centres varied in their practice patterns non-systematically. We checked for this by doing a test of heterogeneity, which asks “Are the differences between the three groups similar between all HSA centres?” Another way of saying this is, “Does the exposed group do better than the comparisons at some centres and worse than the comparisons at other HSA centres?” When the test of heterogeneity is statistically significant, it is an indication that between-centre variation is non-systematic and cannot be adjusted statistically. When this occurred, no other analyses were undertaken since the data were judged too variable to be reliable.

2) It was also possible that although centres varied in their practice patterns, they did so systematically. By including centre as a covariate in the full multiple-regression model, we adjusted for this statistically. Additionally, to see whether the inclusion of centre influenced the measures of association between group and outcome, all models were tested excluding centre. Where an effect was observed, this was reported in the text. Otherwise all multivariate models for health examination outcomes included HSA centre.

For continuous outcomes, analysis of variance (ANOVA) was used to assess heterogeneity of the relationship between exposure group and outcome across centres. A model which included exposure group, HSA centre and the interaction between group and centre was obtained. Significant heterogeneity was considered to be present if the p-value for the group by centre interaction term was statistically significant – i.e. $p \leq 0.05$. For dichotomous outcomes, tables of outcome by exposure group were stratified by HSA centre and the Breslow-Day test for homogeneity of odds ratio across centres determined. This was done separately for each comparison group compared to the exposed group, as only two groups can be considered at any one time for this test. Again a significance level of 0.05 was used for these analyses. Heterogeneity was only assessed for outcomes obtained from the health examinations; this issue was not relevant for Postal Questionnaire items. Results of the tests for heterogeneity are reported in the text of the chapters.

6.5 Descriptive analyses

For each outcome of interest, a descriptive table was included which provided summary information by exposure group (if there was no heterogeneity). For continuous outcomes the mean, standard deviation, median, first and third quartiles were provided for each exposure group. For categorical outcomes, the number and percent of participants in each category, by exposure group, were provided. For variables with more than two levels, the distribution across all categories was shown in the descriptive table, although regression analyses were only undertaken with categories collapsed to dichotomous outcomes. The descriptive analysis was reported in the first of two tables for each outcome.

6.6 Primary analysis

For each outcome, regression models were obtained for Amberley versus exposed and Richmond versus exposed, i.e. the exposed formed the reference group for comparison. Linear regression was used for continuous outcomes, and logistic regression was used for dichotomous outcomes. Potential confounding variables were included in the regression models; age, rank category and posting period were included as covariates in all analyses, whereas other potential confounders such as smoking and alcohol were included, as appropriate, based on the known biology of each outcome. Gender is generally considered a potential confounder for most health outcomes. However, because the number of females involved in DSRS and therefore included in the General Health and Medical Study was small, all females were included in analyses but gender was not considered a potential confounder. Centre was included as a covariate for all outcomes obtained from the health and medical examinations performed at HSA centres. Chapter 8 provides details of all main potential confounders considered for the study. The full list of potential confounders included in analyses for each outcome are outlined in each relevant chapter.

Regression models are reported in three ways:

1) *As the full model.* Here all variables were retained regardless of their significance. The odds ratio or regression coefficient for the group effect was reported in the first line of the second table for each outcome, with all other odds ratios/coefficients reported in the appendices for the relevant chapter. Primary analyses involved obtaining a model which included all covariates of interest. The PROC GLM procedure in SAS was used

for continuous outcomes, and the PROC LOGISTIC procedure was used for dichotomous outcomes.

2) *As the parsimonious model.* Here non-significant variables not associated with the outcome were dropped sequentially from the full model until only significant variables remained. This backward stepwise regression used the model option selection=stepwise. These results are presented in the appendices for each outcome.

3) *Adjusted for "clustering".* Given that participants were examined by particular teams at particular HSA centres, there may be a clustering effect (i.e. the variance for all participants at a centre may be artificially low compared to what might have been if all measurements had been truly independent). Another way of saying this is that people may appear similar simply because they were examined by the same team at the same centre. This clustering effect may be adjusted for by calculating robust variance estimates. Although regression analyses were only performed on outcomes where there was no significant heterogeneity of the group–outcome relationship across centres, and inclusion of centre as a covariate adjusts for between-centre variability, use of robust standard errors would account for any clustering of results by HSA centre. This might reflect differences in measurement applications, for example, where differences between measures within HSA centres tended to be smaller than differences in these measures between HSA centres. Put another way, clustering would denote a lack of independence of observations and was addressed using robust standard errors. Models using these robust standard errors were termed "robust models". Robust standard errors were obtained by using the SAS procedure PROC GENMOD with the repeated statement referring to HSA centre. These are reported in the appendices for each outcome.

6.7 Secondary analyses

Three secondary analyses were undertaken:

1) The first subgroup analyses included all comparison participants but only exposed participants who reported involvement in Program 1.

2) The second subgroup analyses involved all comparisons but only exposed individuals reporting participation in Program 2.

3) The third subgroup analyses was a dose-response analysis, which included all comparison participants and three dose groups for exposed participants (irrespective of which program/s they reported involvement in). Dose was grouped into tertiles based on total length of time of involvement in all programs reported by participants. The three categories were: mild exposure (up to 9 months), moderate exposure (10 to 29 months) or prolonged exposure (30 months or more) (as described in Chapter 4).

As the information on exposure was obtained from the Exposure Questionnaire, exposed participants who did not complete this questionnaire could not be included in the secondary analyses. Due to the number of activities and chemicals within each of the DSRS programs, and in keeping with the overall research questions for the study, no attempt was made to assess or attribute exposure to individual components of DSRS. Although there were seven possible activities associated with DSRS – Program 1, Program 2, Wing Program, Spray Seal Program, storage, mixing and disposal – separate subgroup analyses were only conducted for Program 1 and Program 2. This was because, firstly, these programs had adequate numbers for statistical analyses, and secondly, most individuals involved in other programs or activities were also involved in Program 1 or Program 2. Chapter 4 (Exposure) provides a more detailed discussion of these issues and also provides detail on the identification and classification of the three dose categories. The subgroup analyses involved full models only; the parsimonious and robust models were not obtained for these analyses.

6.8 General principles

For continuous outcomes, statistical significance of covariates was determined by the Type III sums of squares. For dichotomous outcomes, a global Wald test was used to assess statistical significance of covariates. For all models, exposure group is included as categorical variable, with the exposed group as the reference group. Measures of effect for the relationship between exposure status and outcome are in one of two forms. For continuous outcomes, the effect is the difference in (adjusted) mean outcome between the comparison group and the exposed group. One effect measure is provided for each of the comparison groups relative to the exposed group. A negative number indicates that the comparison group has a lower mean outcome than the DSRS exposed group. A positive

value indicates that the comparison group has a higher mean than the exposed group. A value of “0” indicates no differences between exposed and comparison groups.

For dichotomous outcomes, the appropriate measure of effect is the odds ratio. Odds ratios are presented for each of the comparison groups relative to the DSRS group. An odds ratio which is greater than one indicates that the comparison group has higher odds of the outcome than the exposed group. An odds ratio between zero and less-than-one indicates that the odds of the outcome are lower for the comparison group than for the DSRS group. An odds ratio of one means that the odds are equal for the comparison and exposed groups. Although it may seem counter-intuitive to use the DSRS group as the comparator, this was necessary for SAS programming. In order to avoid the potential for transcription errors, this format has been preserved in the Results section of each chapter. However, in the Discussion section of each chapter, the odds ratios have been reversed to provide lay-readers with a more intuitive explanation (i.e. an odds ratio greater than one indicates a greater probability of having a particular outcome in the exposed group). For dose-response analysis, dose was included as a categorical variable, with the “no exposure” group as the reference category, rather than as a continuous variable. While the latter would provide a test for linear trend, there was no evidence that the relationship between increasing level of dose and outcome (or log odds of outcome for dichotomous variables) was linear.

6.9 Summary of presentation

The main body of this report includes two tables for each outcome: the first gives the descriptive analysis; the second gives a table of regression results showing the relationship between exposure group and outcome for the full model for the primary analysis and the three secondary analyses (Program 1, Program 2 and dose-response). The second table provides: the point-estimate for the relevant measure of effect for each of the comparison groups relative to the DSRS group; the 95% Confidence Interval for the measure of effect; and a global test for the statistical significance of the group effect. The tables for continuous outcomes include the R^2 value, which is the proportion of variance explained by the model. This value ranges from 0 to 1, with a higher value indicating a better model. There is no equivalent of this measure for dichotomous regression. The measures of effect and 95% Confidence Intervals for all variables are included in appendices for all of the six models: primary analyses for full, parsimonious, and robust models; and secondary analyses for Program 1, Program 2 and dose-response in that order.

This page has been left blank intentionally
for formatting purposes

7 Quality Assurance and Ethics

Chapter summary

The aim of quality monitoring within the General Health and Medical Study was to facilitate the accurate collection and entry of Health Examination and Postal Questionnaire data. This chapter provides details of quality assurance measures such as training Health Services Australia staff in order to standardise health examinations; developing specific databases for entering and storing study data; including logic checks to identify out-of-range values and to alert the study team to missing data; and establishing procedures for transferring and backing-up study data. Details are also provided about ethical issues such as confidentiality and duty of care to study participants and research staff.

Chapter contents

7.1 Introduction

Quality assurance procedures were important for the optimal conduct of the SHOAMP General Health and Medical Study (the Health Study). Quality control was instituted for each step of the study including the collecting, recording, transferring, entering and checking of data.

7.2 Study structure

Each organisation involved in collecting and recording data for the Health Study (i.e. TUNRA Ltd and Health Services Australia) signed a Service Agreement with the Department of Veterans' Affairs (DVA). The Agreement document described the terms and conditions for collaborating agencies, to ensure that the objectives and milestones of the health study were acknowledged and reached. Regular scheduled meetings of the Scientific Advisory Committee (SAC) and the Consultative Forum provided opportunities for methodical and practical feedback on study protocols, documentation, and issues pertaining to the conduct of health examinations, data analysis and information dissemination.

7.3 Data collection

7.3.1 Postal Questionnaire

The SHOAMP 1800 free-call number connected directly with the office of the study Project Manager (from TUNRA). This facilitated a prompt response to any queries from participants or potential participants regarding the health examination or questionnaire components of the study. In the absence of the Project Manager, the 1800 number was staffed by another member of the research team, and all records of telephone calls were kept locked in the Project Manager's office.

7.3.2 HSA health examination

Prior to the conduct of any health examinations, training sessions for HSA clinical and administrative staff members were conducted with the assistance of TUNRA investigators

and consultants, HSA professionals, and experts in the areas of colour vision, olfaction testing, and breast examination. The 1-2 day training workshop involved tailored education and training and covered all health examination techniques to be performed during the SHOAMP test period. The workshops were based on a series of Instruction Manuals developed for each professional group involved in the assessments: there was an Instruction Manual for nurses, doctors, psychologists and data entry staff (see Appendices 7A, 7B, 7C and 7D). Each manual contained information about SHOAMP, instructions on the conduct of the different health and medical aspects of the examination, checklists, and contact numbers for members of the study team. A copy of a Health Examination Booklet (see Appendix 2N) was given to each health professional, so that they could see where and how to record study data. Training sessions were conducted as close as possible to the starting date for examinations so that the information would remain fresh to all personnel. Ongoing “refresher” visits were conducted as necessary by members of the study team, and phone calls from the Project Manager or Data Manager were made in response to queries from HSA staff about data collection methods, codes used when recording, and the storage of electronic data. At each centre, the equipment used in training was the same as that used during the study proper.

On the first day a centre began the health examinations, a member of the research team visited with HSA staff (the Project Manager and/or Data Manager) to ensure that they were familiar with the tasks required, comfortable with their responsibilities, and knew how to check through each person’s medical booklet.

Regular communication between health examination staff and the research team was encouraged by phone, fax and/or email. Nominated staff at each centre collected the Postal Questionnaire/Consent Form from participants when they arrived for their health examination. Each participant had their information entered for the first time at HSA, and any areas of missing data were resolved before they left. HSA centre staff followed weekly back-up procedures: all psychological, Postal Questionnaire and respiratory data were burnt to a CD and sent to the research team using express postage satchels.

7.3.3 HSA data entry

HSA front desk staff met SHOAMP participants, alerted data entry personnel to their arrival, and confirmed that an Informed Consent Statement had been completed. It was the responsibility of data entry personnel to ensure that while each participant was having the

health examination, their Postal Questionnaire data was entered into the specially developed ACCESS database. Each questionnaire was entered according to the person's unique ID number, which had been pre-printed on the questionnaire before it was posted out by DVA as part of the invitation package. Only valid ID numbers were accepted by the data entry program. Missing or ambiguous information was entered temporarily using "missing" codes until it could be checked with the participant before they left the HSA centre; corrected or additional data were then entered into the database. Data entry staff were responsible for recording in a separate journal any areas of confusion about participant responses or any problems during data entry which meant they had to enter information other than that recorded by the participant. Journal entries were regularly reviewed in conjunction with the data and questionnaire.

Data entry personnel were given a separate Back-Up Manual; updates were provided to each centre during the course of the study. Regular phone calls from the Data Manager (at TUNRA) to each member of the data entry team (at HSA) also assisted with the implementation of any new protocols developed to back up and/or download study data.

7.4 Round One trial phase

A pilot phase (Round One) was conducted with a sub-sample from the exposed group and Amberley comparison group (see Chapter 3), in order to:

- evaluate mail-out materials and data-recording proformas
- evaluate the process of returning questionnaires by post to the project team
- evaluate the use of the ADVISOR system (see Glossary) by DVA for making and recording contact with study participants
- evaluate the use of the ELMA system (see Glossary) by DVA for making participant health examination appointments
- assess the health and general medical assessment procedures administered by HSA personnel
- assess the data entry and back-up protocols
- estimate study consent rates.

Following the Round One trial phase, minor improvements were made to the Postal Questionnaire, including the addition of a checklist of aircraft maintenance tasks specific to DSRS and the aircraft on which it had been conducted. In terms of the health examination, it

was decided to exchange some tasks between HSA medical and nursing personnel to equalise the time allocation between professionals, hence reducing waiting time for participants.

7.5 Data transfer and back-up

Completed Health Examination Booklets and Postal Questionnaires were returned by HSA to the TUNRA study team by priority overnight courier. A fax was sent by HSA when a courier pick-up had been arranged, to inform the team of the number of booklets being sent and whether back-up disks were also included. If the fax differed from the material actually received, a check was made by telephone with HSA staff to see where the error lay and to ensure that no data were missing. When Health Examination Booklets and Postal Questionnaires were returned to the research team, each was logged by hand into a journal and electronically into an Excel spreadsheet.

The journal kept a record of items returned by each study participant and also noted where further follow-up was required. The spreadsheet provided a second record of the type of information that had been returned against each ID number. Data logged on receipt were:

- the Postal Questionnaire
- the Consent form
- the Health Examination Booklet, including results that were stapled into each booklet
  - Spirometry print outs
  - all neuropsychological print outs
  - General Health Questionnaire
  - WAIS booklet
- Faecal Occult Blood Test (FOBT) result card (posted in by participant)
- pathology results (posted in by HSA medical staff after review)
- the Exposure Questionnaire
- Doctor Nomination Form (posted in by participant so that their summary health examination report could also go to their doctor).

Weekly checks were made of the spreadsheet to identify ID numbers where one type of information had been returned without another, and to update the system when new information had been received. For example, if a Health Examination Booklet was returned by HSA, but no Postal Questionnaire, then follow-up with HSA and a search of existing

booklets was conducted to ensure that it had not been misfiled. Ongoing checks of all questionnaire types meant that booklets were matched together, entered into the correct database, and any areas of missing data solved promptly. Regular lists of HSA health examination appointments for SHOAMP were emailed to the Project Manager. These were also cross-checked against returned Health Examination Booklets to ensure that all data collected had been returned to the project team. Again, immediate follow-up with HSA occurred when differences were identified.

7.5.1 Health Service Australia back-ups

At the end of each day the data entry person at HSA had to complete the back-up procedure. The back-up consisted of the Postal Questionnaires entered that day only. The data was written into several text files and a Microsoft Excel spreadsheet. These were saved to a special folder on the data entry computer, ready for the end of week back-up.

7.5.1.1 Spirometry

The Office Medic program, which collected the Spirometry data, was backed up at the beginning of each day. The database containing all the data was copied to the HSA centre server, from which it was collected at the end of the week for the back-up.

7.5.1.2 CIDI and Neurasthenia

Once a week the data entry person went to the program folders for both the CIDI and neurasthenia programs and copied the folder named for the current month, placed it into a specially-designated folder on the data entry computer and then into another folder which identified it as either CIDI or neurasthenia data. All care was taken not to mix these files, as output files for each program had exactly the same file names for each ID number.

At the end of every week, each HSA centre was required to send a back-up CD of all the data collected for that week. This data consisted of files created from the two psychology programs, CIDI and neurasthenia, the Spirometry database, and back-up files from the Postal Questionnaire database. The CIDI and neurasthenia files were those collected for the current month. The Spirometry database gave results for all participants tested at the HSA

centre since testing began. The Postal Questionnaire database back-up files reflected those Postal Questionnaires entered during the week (since the last back-up).

When back-up CDs were received by the Data Manager, the contents of each was catalogued using Microsoft Excel. This was done by listing the relevant ID numbers for each of the file types that had been received. A further listing was also done on individual files for the CIDI and neurasthenia tests to ensure that complete file sets for each study participant had been received.

Once catalogued, the CD contents were then transferred to a secure folder on the TUNRA server. The Postal Questionnaire files were uploaded into a database consisting exclusively of data entered at HSA centres. The Spirometry database was converted into Microsoft Access format, a new database being created for each CD. The CIDI and neurasthenia files were grouped by ID number into separate folders. Due to the nature of these files all care was taken to ensure CIDI and neurasthenia files for one participant did not overwrite the files for another.

When an HSA centre had finished conducting SHOAMP medical examinations, copies of the CD catalogues were compiled to form a master list of all the data collected from that centre. The list was then imported into Access, where it was grouped by ID number, and a report was generated to list all the files present. In this way missing data were quickly identified and a follow-up was conducted with the centre in question.

Back-up procedures for all SHOAMP data were tailored to meet the needs of the research team and the availability of resources at individual HSA centres (e.g. additional steps for those HSA centres without a server compared to those with a server). One person per centre was tasked with weekly back-up of data, or bi-weekly in the case where a server was not available. Appendix 7E details the types of data and procedures for back-up and return to the research team.

7.5.2 TUNRA research team back-ups

At the end of each day one person from the TUNRA data entry team was nominated to back-up the databases that had been used. The files created from the back-up were then moved from the data entry computers to the server under the same folder as the database. These files were also tagged with the login of the data entry person who performed the back-up.

Back-up of SHOAMP data at the project office was done every Friday. This back-up was done to complement a back-up already carried out on the server by data entry staff at the end of each day. A different back-up was conducted each alternate week. The first week of each fortnight was a back-up of the complete set of files from the SHOAMP folder on the server (including the databases used for data entry). The second week of the fortnight was a back-up of only the databases and associated back-up files and a back-up of the compiled CD files collected from each HSA centre. There were security measures in place to prevent anyone besides the Data Manager from having access.

7.6 Data entry

Study questionnaires and the Health Examination Booklet were designed so they would be easy for study participants and HSA staff to complete and for data entry staff to transfer to electronic format. Developed initially using Access 97 (for Windows 98), the database input screens were designed to mimic data records in layout and format. This meant data entry staff did not have to make decisions about where to enter responses to questions.

Checks were built into each input screen to stop data entry staff moving onto the next screen if an input field had been left empty. For example, if data from an item in the Postal Questionnaire were missing, codes were developed for each of the field types recognised by the database (text, numeric and date fields) so that no fields were left blank. Also, codes for “not applicable” and provision for “skip” questions were also incorporated, to un-enable certain fields where data were not required. For example, if the participant never worked on aircraft, a “no” entry into the database would produce an automatic-fill response of all “aircraft” fields in that input screen.

Each input screen had an accompanying “help screen”, which provided summary explanations of the answers required for each field. Focused primarily on the numeric fields, the “help screen” indicated the correct value to be entered for each response. Each screen also contained a “special codes” feature, which detailed the exact codes to use for “empty”, “refused” or “not applicable” response options. Each member of the data entry team was required to use their own unique ID number to log into the database.

All data were double entered. The Postal Questionnaire was entered the first time by HSA staff, if the participant took their booklet along to the health examination, and a second time by TUNRA data entry staff. A number of questionnaires was also returned by post directly to

the research team. These were entered twice by different members of the study data entry team and filed separately in anticipation of a Health Examination Booklet being sent in as well. All other data were entered twice by different personnel from the research team. A separate team of six data entry personnel, trained in the use of the Australian Standard Classification Occupation (ASCO) and the Australian New Zealand Standard Classification of Industry (ANZSIC) coding processes, coded and entered all civilian occupations and industry types. The training program consisted of lessons and exercises produced by the Australian Bureau of Statistics (ABS). All ASCO and ANZSIC data were coded twice, with the second coder blinded to the first result. Agreement between the two coders was then checked, and where an occupation or industry had been coded differently, a member of the study team re-coded the information.

7.7 Data checking

Various logic checks were built into the data entry process. Checks were programmed into the database to ensure that all ID numbers were valid and were being used correctly for either male- or female-specific entry. If the study number was not valid, then it was rejected and the data entry person could not continue. Although the databases contained numeric oth and text fields, the primary focus for range checks was the numeric fields. These fields required a value from a pre-determined range of values to be entered. If the input value was not within the given range, then a message box appeared restating the range required. To further facilitate correctness of data entry, the range of values for each on-screen questionnaire item was provided next to the field that required it. Where one response was required from a multitude of options, the system would reject data entry outside those options. Further, where the response to an initial question impacted on subsequent answers (i.e. record of alcohol habits over time), if discrepant values were given, the most extreme value was chosen. For example, if a participant indicated that they drank lightly and also drank heavily, then as the logic check would not allow varying degrees of the same answer, the higher degree would be taken, which in this case would be “drinking heavily”. Formal data checking procedures were applied to four sources of data: the Exposure Questionnaire, Postal Questionnaire, Health Examination Booklet and the Female Reproductive Questionnaire. In addition to general checks for areas of missing data in all numeric fields in all questionnaires, specific checks were conducted on:

- lipoma (part of skin examination by doctor)
- psychological test scores (as calculated by psychologist)

- Mini Mental Status Examination score (doctor administered)
- contact details for participant and nominated doctor (self-reported by participant in the Postal Questionnaire)
- medications (also self-reported by participant in Postal Questionnaire).

First and second entry data were electronically compared using PROC COMPARE, a procedure of the SAS system software (SAS V8.2, SAS Institute Inc., Cary, NC, USA). Any differences detected were checked against the paper record and corrected where appropriate. The same checks were re-run and re-corrected where necessary, until no differences were detected. Four SAS programs were created for the purpose of comparing all numeric fields of first and second entry data. Apart from contact details provided in the Postal Questionnaire for the participant and their nominated doctor and self-reported medications, only numeric fields were checked electronically. Contact details were used to mail out participant's results from the health check and so had to be electronically perfect, as did medications that were linked to the summary report of results.

Each program for comparing database entries had a similar format. All data from the Exposure Questionnaire were stored in a single table in the two Access databases. The checking program imported that table from the first and second entry databases and stored these as two data sets in SAS. A selected set of ID numbers identifying records to be compared were also imported, and records not required were dropped from each data set. A comparison of all variables in each data set was then undertaken. Discrepancies were highlighted and the paper record was used to amend the data in the first, second or both databases. The programs for the Postal Questionnaire and the Health Examination Booklet were more complicated because of multiple tables. Additionally, there were several multiple entry tables. These were compared using JMP statistical software (JMP V4.05, SAS Institute Inc., Cary, NC, USA), as this process required some manual manipulation of tables. Text fields were, for the most part, handled manually. The Postal Questionnaire contained a large number of text fields and the Exposure Questionnaire contained a small number of large text fields. Apart from contact details and medications (which were not coded), manual checking of all text fields occurred when they were processed for data analysis.

In terms of checks for lipoma data, comparisons were made between the first and second data entry in the Health Examination Booklet medical database. Checks were made, for example, to determine whether lipomata were present or absent, and if present, to make sure both entries recorded the same information for size (length and width) and number of lipoma. Missing data codes were checked for all psychological tests, and where possible,

scores were recalculated from individual items for completeness. The Mini Mental Status examination score was also checked and recalculated. Contact details were checked according to a specific set of “rules” developed to standardise the way in which personal and address information were entered. Medications were checked for errors in spelling and medical correctness, as these data were linked to the participant and medical practitioner summary reports of health examination results and had to reflect true medications for the information to be useful.

Separate checks were carried out on codes allocated to self-reported occupation and industry details provided by participants in Section 7a of the Postal Questionnaire (PQ). Each person was instructed to record details of any employment held outside service in the Defence Force (i.e. civilian employment), so that this information could help to inform the assessment of exposure to solvents and other relevant chemicals outside Defence Force activities. This was considered particularly important if a respondent had reported relevant exposures since being involved in F-111 DSRS activities or during the period prior to their SHOAMP health examination when potential effects to some health outcomes may occur. Random samples of completed PQ data were reviewed to ensure that the correct code had been allocated to each occupation and industry the participant recorded. Incorrect entries were reviewed and amended, and further checks were carried out on the codes entered by individual data entry personnel if repeated inaccuracies were identified. It was also found that some respondents had reported military occupations in the civilian section of the PQ. Therefore, prior to analysis of civilian exposures, the following steps were taken:

- cross-checking all jobs coded with an industry code of “8200 defence” to ensure validity of civilian status
- checking whether respondents with a rank code of “civilian” had reported any military tasks in PQ Section 7B
- checking whether respondents who were civilian only, were reporting their DSRS activities in PQ Section 7A; this exposure could not be assessed as a potential confounder as it actually formed part of their exposure status
- checking whether those respondents who had returned an empty civilian section of their questionnaire had reported a military history; if both were returned empty the participant was excluded from this analysis
- checking for jobs reported during the time period of their involvement in a DSRS program
- checking for any civilian jobs reported in PQ Section 7B of the questionnaire, i.e. civilian jobs reported in the military section.

7.8 Final data accuracy

Prior to analyses, all outcome data were checked one final time to ensure that data included in analyses were accurate, that all results had been scored where possible, and that areas of missing data had been minimised. Each data entry database had initially been programmed with “normal” ranges for each variable type, outside of which an “abnormal” indicator would alert the study team to data variability. During the final data checking process, a number of inconsistencies were detected and corrected:

a) *Ishihara colour vision testing.* It turned out that different HSA centres had used different chromatic plate sets, so these results had to be adjusted.

b) *Visual acuity testing.* The data checks revealed that different variations of the Snellen charts were being used and at different distances away from the participant. Therefore, the pre-programmed “normal” ranges of good vision were recording “abnormal” results (based on the anticipation that each centre used the same testing chart/procedure), and individual codes had to be re-programmed per centre to account for the different charts.

c) *Post-Ventolin® spirometry testing.* During health examination data collection, the checking process identified that one HSA centre had a significantly lower number of post-Ventolin® lung function results (33% non-completion vs 2.1-8.4% non-completion rate at other centres). This information was fed back to HSA for further investigation.

d) *Olfaction.* Similarly, a large number of missing values and missing overall score were noticed for the Sniffin’ Sticks test. Checks of individual Health Examination Booklets revealed clinician differences in the way results were recorded. As they did not have a “normal” configuration, each had to be checked with the attending HSA doctor, re-scored accordingly and a total calculated.

e) *Mini Mental Status Examination (MMSE).* Checking identified a substantial number of missing values throughout the MMSE for individual items and for overall score. Re-scoring of individual health examination results for MMSE was necessary where a clinician had not written an individual result or overall score. This was done by a study consultant experienced in the administration and scoring of MMSE and who was “blinded” to the exposure status of participants. Some items had to remain missing, as it was not possible to re-score them after the fact (e.g. having the participant fold a piece of paper could not be re-scored by the study team).

7.9 Ethics

7.9.1 Ethics approval

Ethical approval for Phase III of the SHOAMP General Health and Medical Study, was granted from The University of Newcastle Human Research Ethics Committee, the Australian Defence Force Human Research Ethics Committee and the Department of Veterans' Affairs Ethics Committee. Letters of approval from these committees are shown in Appendices 7F, 7G and 7H respectively. In accordance with Commonwealth legislation, personal information acquired (or developed) during the study can only be used for the purposes of SHOAMP in order to investigate the relationship between involvement in F-111 DSRS activities and adverse health outcomes.

7.9.2 Consent

Written informed consent was obtained from all participants. Participants were assured that participation in the study was voluntary, that they were free to withdraw at any time during the study, and that non consent would not in any way affect their current or future treatment, career opportunities within the Defence Force (or civilian employment), or Veteran entitlements. Participants were assured of the anonymity of the data, and that individual results would not be made available to any party other than the participant and their nominated local medical centre, should they choose to nominate one. The Informed Consent Statement was designed in such a way that participants could agree to assist with one aspect of the study but not another. The areas for consent were:

- completion of the Postal Questionnaire
- participation in the health examination (with separate consent checks at HSA for blood collection and Ventolin® administration during spirometry testing)
- storage of blood sample
- storage of de-identified data after the study by the Australian Institute of Health and Welfare, with a list of codes held by the Royal Australian Air Force Association, through which personal data could be identified and accessed
- access by researchers to some Australian Defence Force medical, psychological and fitness testing results.

Individuals who did not wish to participate in the SHOAMP health examination were asked to complete the Postal Questionnaire (as well as the Exposure Questionnaire and Female Partner Reproductive Questionnaire as appropriate).

7.9.3 Participant burden and duty of care

It was anticipated that the health and medical assessments conducted by HSA would take approximately four hours for each participant. Evidence from the Gulf War Study indicated that the participants were not overburdened with this time commitment. HSA recognised the need for patient care and follow-up (i.e. should a participant become upset at any stage, trained HSA counsellors were available at all participating centres, and all members of the team were aware that testing would cease if a participant did not wish to continue). Each participant was asked during their assessment if they required a break for refreshments or to move around. Each potential study participant received a list of people for them to contact in case they had any concerns about their participation at any time. Contact details for each Ethics Committee were freely provided to all potential participants, in case they wished to contact an independent person about any part of the study process. While the mail-out questionnaire also appeared lengthy, participants had many days to complete the survey prior to their Health Examination appointment. Studies such as the Gulf War Study had used questionnaires of a similar length, and participants had not found these too difficult or burdensome.

At the time of the health examination, HSA clinicians were able to give a participant an immediate feedback letter (Appendix 7I) when any of the following criteria were identified:

- significantly underweight, particularly if there was a history of recent unexplained weight loss
- poor vision
- blood, protein or glucose detected in the urine test
- irregular pulse, persisting bradycardia or tachycardia
- hypertension
- significantly abnormal spirometry – FEV1 < 1 litre, FEV1/VC < 40%
- abnormal chest sounds (i.e. bronchial breathing, wheezes, crackles or a rub)
- a likely malignant skin lesion
- enlarged lymph nodes
- a breast lump

- significantly abnormal blood test results likely to be due to a pathological process (e.g. anaemia, markedly elevated liver function tests).

Participants were given a summary feedback report of the results of their health examination (Appendix 2J). A copy of the same results was also forwarded to a nominated medical practitioner, where the participant had provided consent and contact details. If a participant chose not to nominate a medical practitioner, the summary letter and health examination results were given to the participant only, with a "Doctor Nomination Form" enclosed so they could nominate a general practitioner to receive a copy of the results at a later date. Serving Defence Force members who did not have a personal medical practitioner outside the Defence Force could nominate a local or otherwise convenient medical practitioner to whom their test results could be given.

7.9.4 Confidentiality

As SHOAMP was commissioned by the Royal Australian Air Force and was conducted in collaboration with the Department of Veterans' Affairs, it was anticipated that a number of participants might have concerns about the independence of the research team and the confidentiality of their health information. Participants might, for example, be concerned that sensitive information disclosed during their general health and medical examination would be made available to either DVA or Defence. To ensure the confidentiality of participants' information, a number of precautionary measures were put in place as part of the SHOAMP quality assurance procedures:

1) Each participant was allocated a unique ID number, against which their results were recorded.

2) Any documents containing a participant's details were stored separately from their medical examination results in locked filing cabinets.

3) All information was reported as de-identified group data with no individual able to be identified.

4) Individual records are subject to the provisions of the *Privacy Act 1988* which regulate their use, storage and disclosure. No third party will be provided with any participant's results without the express written permission of that participant to the research team and without the third party being identified.

5) No summary report of participant results was, or will be, forwarded to a general practitioner without the express written consent of that participant. However, examining

HSA medical personnel had a duty of care to ensure that participants were appropriately referred if a medical condition requiring further follow-up was discovered during the study.

6) It is the responsibility of any SHOAMP participant currently serving in the Australian Defence Force to inform their employer of the presence of any illness or condition that could compromise their operational abilities. However, under no circumstances did, or will, any member of the research team or HSA provide such information to the Departments of Defence or Veterans' Affairs without the express written consent of the participant.

Requests for participant information can come from the participant themselves, if they want copies of their health examination information, or from a third party (medical practitioner, specialist or legal counsel) who wants copies of the information for referral or compensation purposes. No release to any third party was, or will be, approved without the express written permission of the participant who originally provided the information. Although each request has been treated separately on its merits, a standardised form (Appendix 7J) was developed to facilitate equal treatment of requests received and the prompt delivery of documents where the request was legitimate. Also, it was not uncommon for study participants to return, along with their questionnaires, *additional* information about their health condition and/or work experiences. To protect the confidentiality of this information, all attachments were stored separately from returned questionnaires. Any identifying details for a participant were removed from the Postal Questionnaire (i.e. Informed Consent Statement), logged, then stored separately away from their health information.

7.10 Access to data

Information managed by the DVA about individuals (RAAF personnel and civilians) known or thought to have had some involvement with the DSRS process was only used by DVA staff employed specifically to work on participant contact and recruitment and was not available for any other purposes within DVA. DVA staff, as part of their employment conditions, are bound by the rules governing confidentiality of information. They are also bound to operate within the restrictions imposed by the Information Privacy Principles. Prior to the receipt of consent information, DVA had sole access to the information of potential participants that was used during the recruitment process. However, DVA did not have access to any health or medical data obtained as a result of the General Health and Medical Study. The study

team had access to the de-identified information for the purposes of data management and analysis. Analysis was conducted on records identifiable only by a specially allocated ID number. Access to de-identified data was limited to certain members of the research team, and all members of the team were required to sign a confidentiality agreement against disclosure of personal information of participants (Appendix 7K).

7.10.1 Data storage

Access to information obtained during the study was by password, only available to staff who had signed a confidentiality agreement. Back-up copies of all electronic data were securely stored and contained no information that permitted any individual to be identified. Data in hard copy and electronic form will be kept at The University of Newcastle for a minimum period of 15 years.* At the end of the study, the file of all Postal Questionnaire and medical test data will be stored in a de-identified format by the Australian Institute of Health and Welfare. A separate de-coding “key” will be safeguarded by the National President and National Secretary of the Royal Australian Air Force Association. This file will only be able to be accessed if a participant decides at some time that they want access to their own personal information, or if the research team finds some information they consider important enough to inform the participant of.

* As determined by entry 20.5.2 of the General Disposal Authority University Records, issued by State Records under the State Records ACT 1998, research data involving human subjects and potential long term effects must be retained for a minimum of 15 years after action is completed. This retention period has been based on the recommendations in the joint statement of the National Health and Medical Research Council (NHMRC) and Australian Vice-Chancellors’ Committee (AVCC).

8 Potential Confounders

Chapter summary

This chapter describes the main potential confounders that were included in the General Health and Medical Study analyses. Essential (potential) confounders used in every analysis were age, rank category and posting period. Key (potential) confounders used in some analyses as appropriate included body mass index, alcohol and smoking behaviours, diabetes, level of education, depression and anxiety. Descriptive statistics indicate imbalances in many of these factors, and this supports the need to adjust for these in subsequent analyses.

Chapter contents

8.1 Introduction

This chapter describes each essential and key potential confounder included in the Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP) health study data analyses. For each potential confounder, the method of measurement and the distribution across all three groups (exposed and two comparisons) are explained.

There were three levels of potential confounders, depending upon their application during each primary outcome analysis:

- Essential potential confounders were age, rank and posting; these were included in every analysis. As discussed in Chapter 6 (Analysis), HSA centre was also included when the outcome was measured as part of the health examination.
- Key potential confounders included height, weight, body mass index (BMI), alcohol, smoking, diabetes, education level, and civilian exposures; these were included in some analyses when biologically indicated.
- Outcome-specific potential confounders included medications (for cardiovascular disease), caffeine (for female partner reproduction), or visual acuity (for colour vision); these were included in analyses of only a few particular outcomes with specific indications.

Table 8.1 lists the potential confounders used for each outcome domain.

In the following sections the method of measurement for each potential confounder and its distribution across the three groups are detailed. Analysis tables include all categories of variables as well as missing values. Some categories have been collapsed where numbers in cell sizes are small and where the combining of categories was deemed biologically/clinically sensible. Distributions of potential confounders were compared between the three groups using the chi-square statistic for collapsed variables (used in analyses of outcomes) and excluding missing values. Analysis of continuous variables required that the distributions were normally distributed and that other modelling assumptions were satisfied.

Table 8.1 : Potential confounders for each Health Study outcome

Outcome	Full model adjusted for…
General health and well-being	Age + Rank + Posting + Smoking + Alcohol + BMI for hospital admissions + Smoking + Alcohol + Education + Civilian Solvent Exposure for SF-36 quality of life
Cardiovascular health	Age + Rank + Posting + Smoking + Alcohol + Civilian Solvent Exposure + Beta Blocker + Nitrate + Aspirin + ACE Inhibitor use + HSA centre
Respiratory health	Age + Rank + Posting + Smoking + Civilian Solvent Exposure + HSA centre for airways disease outcomes
Dermatological and breast abnormalities	Age + Rank + Posting + Smoking for melanoma and skin cancer + HSA centre + Smoking + Alcohol + BMI for breast abnormalities
Neurological outcomes	Age + Rank + Posting + Smoking + Civilian Solvent Exposure for olfaction analysis + Diabetes for colour vision and peripheral neuropathy + Visual Acuity for colour vision + Height + Alcohol + Civilian Solvent and Lead Exposure for peripheral neuropathy + HSA centre for health examination outcomes
Sexual function and reproductive health	Age + Rank + Posting + Civilian Solvent and Lead Exposure + Smoking + BMI + Depression + Anxiety for male sexual function + Age at pregnancy, Rank and Posting Category of male partner + Smoking + Alcohol + Caffeine for female reproductive health
Mental health	Age + Rank + Posting + Alcohol + Smoking + Education + Civilian Solvent Exposure + HSA centre
Neuropsychological outcomes	Age + Rank + Posting + Alcohol + Smoking + Education + Civilian Solvent Exposure + Depression + Anxiety + HSA centre

8.2 Measurement and distribution of essential potential confounders

8.2.1 Age

In most studies, age is viewed as a potential confounder. Age is generally strongly associated with end points of interest in occupational epidemiology such as diseases, physiological characteristics, and so on. For the SHOAMP General Health and Medical Study, the exposure of interest is time-related: duration of employment, time since first DSRS exposure, and cumulative exposure, could all imply that the older an individual, the greater the potential for exposure. There were more exposed participants in the younger two and oldest age categories, whereas there were more middle-aged respondents among the two comparison groups (Table 8.2), and this difference was statistically significant ($\chi^2_{14} = 28.81, p = 0.011$). However, when considered as a continuous variable, there was no difference on average between ages among the three groups, p=0.23 (Table 8.3).

Table 8.2 : Distribution of essential potential confounders – Age categorical

	Amberley		Richmond		Exposed	
	N	***%***	**N**	***%***	**N**	***%***
Total	**406**	***100***	**516**	***100***	**616**	***100***
AGE						
≤ 29yrs	14	*3.5*	17	*3.3*	33	*5.4*
30-34yrs	41	*10*	48	*9.3*	81	*13*
35-39yrs	77	*19*	75	*15*	106	*17*
40-44yrs	107	*26*	134	*26*	125	*20*
45-49yrs	86	*21*	125	*24*	121	*20*
50-54yrs	42	*10*	68	*13*	68	*11*
55-59yrs	30	*7.4*	35	*6.8*	50	*8.1*
≥ 60yrs	9	*2.2*	14	*2.7*	32	*5.2*

Table 8.3 : Distribution of essential potential confounders – Age continuous

	Amberley	Richmond	Exposed
Number of observations	**406**	**516**	**616**
AGE (YEARS)			
Mean	43.9	44.8	44.1
Standard Deviation	7.8	8.0	9.3
Median	43.4	44.3	43.5
Minimum	26.0	25.4	24.4
Maximum	68.0	73.3	73.3
Lower 95%	43.1	44.1	43.3
Upper 95%	44.6	45.5	44.8
Number missing	0	0	0

8.2.2 Posting

Posting provides an indicator of time since exposure (for exposed individuals) and also possibly the type of exposure (given that certain work activities were carried out depending upon the posting period). Posting was classified into one of five time periods: 1975-79, 1980-84, 1985-89, 1990-94, 1995-99 (Table 8.4).

Table 8.4 : Distribution of essential potential confounders – Posting on exposure category

	Amberley		Richmond		Exposed	
	N	***%***	**N**	***%***	**N**	***%***
Total	**406**	***100***	**516**	***100***	**616**	***100***
POSTING						
1975-1979	135	*33*	188	*36*	215	*35*
1980-1984	102	*25*	131	*25*	132	*21*
1985-1989	94	*23*	123	*24*	129	*21*
1990-1994	46	*11*	47	*9.1*	94	*15*
1995-1999	29	*7.1*	27	*5.2*	46	*7.5*

Posting period was different between the three groups, with fewer exposed from the posting period 1980-1984 where there were 21% in the exposed group and 25% in each of the comparison groups, and proportionally more exposed from the posting period 1990-1994 where there were 15% compared to 9.1% from Richmond and 11% from Amberley (Table 8.4). The difference overall was borderline to significant ($\chi^2_8 = 15.50, p = 0.050$).

8.2.3 Rank

The inclusion of rank as a potential confounder provides a measure of the socio-economic status of Health Study participants. Rank was classified into three categories (as obtained from the AFPEMS database): (a) Enlisted, (b) Non-Commissioned Officer and (c) Officer. Among the exposed group there were 38 civilians and 22 of unknown rank; however, all rank categories were known for the two comparison groups and there were no civilians in these groups (Table 8.5). Comparison of rank across exposure groups required dropping 60 exposed who were civilian or for whom there was no rank code. There was a similar distribution of rank across the three groups ($\chi^2_4 = 1.01, p = 0.91$).

Table 8.5 : Distribution of essential potential confounders - Rank

	Amberley		Richmond		Exposed	
	N	***%***	**N**	***%***	**N**	***%***
Total	**406**	***100***	**516**	***100***	**616**	***100***
RANK						
Civilian	0	*0*	0	*0*	38	*6.2*
Enlisted	261	*64*	321	*62*	359	*58*
Non-Comm. Officer	133	*33*	180	*35*	179	*29*
Officer	12	*2.9*	15	*2.9*	18	*2.9*
Unknown	0	*0*	0	*0*	22	*3.6*

8.2.4 HSA centre

As discussed in the analysis section, HSA centre was routinely included in the full model in order to check and adjust for any variability between centres. Table 8.6 lists the number and proportion of each group examined at each centre. Although there were ten centres, several

shared the same staff and were therefore collapsed together (Townsville into Brisbane, Hobart into Melbourne, Darwin into Adelaide for medical data, and Darwin into Brisbane for psychological data). It is important to note that different proportions of each study group were examined at each centre. For example, 41% of the exposed were examined at Brisbane compared with 35% and 25% of those in the Amberley and Richmond groups respectively, while 31% of the Richmond group were examined at Parramatta compared with 21% and 15% from the Amberley and exposed groups respectively.

Table 8.6 : Number and proportion of each group examined at each HSA centre

	Amberley		Richmond		Exposed	
	N	**%**	**N**	**%**	**N**	**%**
Totals	**406**	***100***	**516**	***100***	**616**	***100***
Adelaide	18	*4.4*	34	*6.6*	22	*3.6*
Brisbane	142	*35*	128	*25*	252	*41*
Canberra	12	*3.0*	28	*5.4*	15	*2.4*
Darwin	8	*2.0*	9	*1.7*	5	*0.81*
Hobart	9	*2.2*	5	*0.97*	2	*0.32*
Ipswich	49	*12*	37	*7.2*	161	*26*
Melbourne	30	*7.4*	33	*6.4*	19	*3.1*
Newcastle	16	*3.9*	46	*8.9*	21	*3.4*
Parramatta	85	*21*	162	*31*	95	*15*
Perth	25	*6.2*	23	*4.5*	16	*2.6*
Townsville	12	*3.0*	11	*2.1*	8	*1.3*

This unbalanced design may lead to difficulty in assigning what proportion of the variance in the full regression model is due to group and what proportion is due to office. For example, given that many of the exposed group were seen at Brisbane HSA, it is difficult to adjust for possible variability at the Brisbane office without (possibly) “adjusting out” some of the group effect. For this reason, we tested each regression model with and without office, as a form of sensitivity analysis. Where this affected the results, it was noted in the text.

8.3 Measurement and distribution of key potential confounders

8.3.1 Body Mass Index

Body Mass Index (BMI) was calculated using the individual height and weight data recorded by HSA nursing staff prior to the conduct of the lung function tests. BMI was calculated as a person's weight (kg) divided by their height squared (metres^2). BMI was analysed as a continuous variable as well as being categorised according to the World Health Organisation recommendations:[1]

- <20 underweight
- 20 – 24 normal weight range
- 25 – 29 overweight
- 30 – 40 obese
- >40 grossly obese.

BMI is reported as a categorical variable in Table 8.7 and as a continuous variable in Table 8.8. Due to small numbers of participants in some BMI categories, "<20 underweight" and "20-24 normal weight range" were combined, as were "30-40 obese" and ">40 grossly obese" in the following analysis and for use as covariates in all multiple regression analyses. According to the categorical data, 2.0% of participants in the exposed group were classified as underweight but only 0.25% and 0.19% in each of the Amberley and Richmond groups respectively.

Table 8.7 : Distribution of key potential confounders – BMI categorical

	Amberley		Richmond		Exposed	
	N	**%**	**N**	**%**	**N**	**%**
Totals	**406**	***100***	**516**	***100***	**616**	***100***
BMI						
Underweight	1	*0.25*	1	*0.19*	12	*2.0*
Normal weight range	60	*15*	87	*17*	126	*20*
Overweight	186	*46*	252	*49*	294	*48*
Obese	147	*36*	166	*32*	180	*29*
Gross obese	12	*3.0*	9	*2.0*	4	*0.65*

Table 8.8 : Distribution of key potential confounders – BMI continuous

	Amberley	Richmond	Exposed
Number of observations	**406**	**516**	**616**
BMI (KG/m2)**			
Mean	29.54	28.75	28.18
Standard Deviation	4.92	4.12	4.26
Median	28.82	28.36	27.78
Minimum	19.29	19.88	18.22
Maximum	59	47.62	50.28
Lower 95%	29.06	28.39	27.84
Upper 95%	30.02	29.11	28.52
Number missing	0	1	0

On the other hand, 3.0% of the Amberley comparison group were classified as grossly obese compared to 1.8% and 0.65% of the Richmond comparison group and the exposed group respectively. There was a larger proportion of those with a normal BMI category among the exposed: 20% compared with 17% and 15% of the Richmond and Amberley groups respectively. There was a larger proportion of obese among the Amberley group: 36% compared with 32% and 29% of the Richmond and the exposed group respectively. The overall difference in proportions with respect to BMI classifications between the three groups is strongly significant, $\chi^2_4 = 14.93, p = 0.005$.

The difference between the means of the three groups is also significant (p<0.0001), where Bonferroni multiple comparison indicates a significant difference between Amberley and Richmond and between Amberley and exposed (Table 8.8). There was no difference in mean BMI between Richmond and exposed.

8.3.2 Alcohol

Alcohol intake was collected via self-reported items in the Postal Questionnaire and was initially grouped into six categories:

- *Teetotaller* – if a respondent indicated there was not a period in their lifetime when they drank alcohol regularly and they had not had a drink in the last three months (Postal Questionnaire [PQ] items 4.2 and 4.3).

- *Safe drinker lifetime* – if a respondent recorded ≤ 4 standard drinks per day in any one cell of PQ item 4.10 [for men] or ≤ 2 standard drinks per day [for women].
- *Moderate drinker* – if a respondent recorded > 4 standard drinks per day in their drinking calendar [for men] or > 2 standard drinks per day [for women], but ≤ 6 standard drinks per day in any cell of PQ item 4.9 [for men] or ≤ 4 standard drinks per day [for women].
- *Hazardous drinker bingeing* – if a respondent indicated they consumed > 6 standard drinks [for men] or > 4 standard drinks [for women] for no more than three days a week (for PQ item 4.9).
- *Hazardous drinker chronic* – if a respondent indicated they consumed > 6 standard drinks [for men] or > 4 standard drinks [for women] for four or more days a week (PQ item 4.9).
- *Former hazardous drinker* – if a respondent answered "yes" to PQ item 2.118 (alcohol abuse or dependency) and answered "more than three months ago" to PQ item 4.1 (when was the last time you had a drink of any kind of alcoholic beverage?).

Self reported drinking behaviour of respondents is presented in Table 8.9.

Table 8.9 : Distribution of key potential confounders – Drinker type

	Amberley		Richmond		Exposed	
	N	**%**	**N**	**%**	**N**	**%**
Totals	**406**	***100***	**516**	***100***	**616**	***100***
DRINKER TYPE						
Former Hazardous Drinker	13	*3.2*	6	*1.2*	21	*3.4*
Hazardous Drinker Bingeing	128	*32*	150	*29*	177	*29*
Hazardous Drinker Chronic	25	*6.2*	31	*6.0*	36	*5.8*
Moderate	62	*15*	109	*21*	139	*23*
Safe Drinker	169	*42*	210	*41*	216	*35*
Teetotaller	1	*0.3*	0	*0*	2	*0.32*
Unknown	8	*2.0*	10	*2.0*	25	*4.1*

There were 42 respondents who did not provide sufficient information to be classified. To analyse the table, “Teetotallers” was combined with the group classified as “Safe Drinkers”, and “Hazardous Drinker Bingeing” was combined with “Hazardous Drinker Chronic”. With respect to drinking behaviour, there was a significant difference detected between the three groups ($\chi^2_6 = 16.86$, $p = 0.010$), with fewer former hazardous drinkers from Richmond (1.2% vs 3.2% and 3.4% from Amberley and the exposed groups respectively) and noticeably fewer moderate drinkers from the Amberley group (15% vs 21% and 23% for Richmond and the exposed group).

8.3.3 Smoking

Smoking is a key determinant of poor health outcomes. Smoking behaviour was grouped into three categories:

- *Ex-smoker* – if a respondent indicated “yes” to PQ item 4.21 (in your entire life have you smoked at least 100 cigarettes?), then “not at all” to PQ item 4.22 (do you now smoke cigarettes everyday, some days or not at all?).
- *Current smoker* – if a respondent indicated “yes” to PQ item 4.21 and either “everyday” or “some days” to PQ item 4.22.
- *Never smoked* – when a “no” response was given for PQ item 4.21.

Self-reported smoking behaviour of respondents is presented in Table 8.10.

Table 8.10 : Distribution of key potential confounders – Smoker type

	Amberley		Richmond		Exposed	
	N	**%**	**N**	**%**	**N**	**%**
Totals	**406**	***100***	**516**	***100***	**616**	***100***
SMOKING CATEGORIES						
Current smoker	98	*24*	71	*14*	120	*19*
Ex smoker	153	*38*	193	*37*	229	*37*
Never smoked	149	*37*	244	*47*	242	*39*
Unknown	6	*1.5*	8	*1.5*	25	*4.1*

There were 39 participants who did not provide information about smoking behaviour. There were noticeably fewer current smokers among the Richmond cohort: 14% compared to 24%

in the Amberley group and 19% in the exposed group. A larger proportion of the Richmond group never smoked: 47% versus 37% and 39% for the Amberley and exposed groups respectively. Overall, the difference in smoking behaviour between the groups was significantly different, $\chi^2_4 = 20.03, p = 0.0005$.

8.3.4 Diabetes

The Postal Questionnaire asked respondents if they had ever been diagnosed by a physician as having "diabetes". Diabetes has been implicated in a variety of adverse health outcomes, particularly when the disease is not well managed. Self-reported prevalence of diabetes among the three groups is presented in Table 8.11. There were no detectable differences between the three groups with respect to diabetes, $\chi^2_2 = 1.36, p = 0.51$. There were 52 participants for whom there is no information on diabetes status.

Table 8.11 : Distribution of key potential confounders – Diabetes

	Amberley		Richmond		Exposed	
	N	**%**	**N**	**%**	**N**	**%**
Totals	**406**	***100***	**516**	***100***	**616**	***100***
DIABETES						
Diabetes	14	*3.5*	14	*2.7*	24	*3.9*
No Diabetes	379	*93*	490	*95*	565	*92*
Unknown	13	*3.2*	12	*2.3*	27	*4.4*

8.3.5 Education

Education reflects socio-economic status. Educational background may also influence psychological testing performance. As part of the Postal Questionnaire, participants were asked to give details of the level of schooling they had reached, which were categorised as:

- primary school only/some high school
- completed high school
- trade/apprenticeship
- certificate/diploma
- university degree, higher university degree, or currently enrolled.

Participants' self-report of highest qualification obtained is presented in Table 8.12. There were 46 participants who did not provide information about educational status. There was a strong significant difference between the three groups with respect to educational attainment, $\chi^2_8 = 303.81, p < 0.0001$. In particular, 48% of the Amberley group had not obtained education beyond high school, compared with 7.6% and 12% of the Richmond and exposed groups respectively.

Table 8.12 : Distribution of key potential confounders – Education

	Amberley		Richmond		Exposed	
	N	**%**	**N**	**%**	**N**	**%**
Totals	**406**	***100***	**516**	***100***	**616**	***100***
EDUCATION						
Primary School or some High School	84	*21*	17	*3.3*	37	*6.0*
Completed High School	109	*27*	22	*4.3*	37	*6.0*
Trade/ Apprenticeship	47	*12*	228	*44*	249	*40*
Certificate/Diploma	108	*27*	175	*34*	206	*33*
University Degree	48	*12*	65	*13*	60	*10*
Unknown	10	*2.5*	9	*1.7*	27	*4.4*

Most of the difference can be explained by the larger proportion of Richmond and exposed participants who reported undertaking or completing a trade or apprenticeship: 44% and 40% respectively compared with only 12% of the Amberley group. Between 10-13% of the three groups had undertaken or were completing a degree or higher degree at university.

An alternative measure of education or pre-morbid functioning was assessed by the National Adult Reading Test (NART).[2] The test comprises a list of 50 words printed in order of increasing difficulty. The participant reads aloud down the list of words, and the number of errors is recorded. The higher the score (i.e. the more errors), the poorer the performance. The NART is reported as a continuous variable and is presented in Table 8.13. Three participants scored the sample minimum of three errors, and one participant scored the test and sample maximum of 50 errors. The mean number of errors varied significantly between

the groups (p=0.03) with the Amberley group scoring a mean of 22 errors, while the Richmond group and the exposed group scored a mean of about 21 errors.

Table 8.13 : Distribution of the NART

	Amberley	Richmond	Exposed
Number of observations	**N=406**	**N=516**	**N=616**
National Adult Reading Test (NART)			
Mean	22.0	20.6	21.2
Standard Deviation	8.7	7.5	8.3
Median	21.0	20.0	20.0
Minimum	3	3	3
Maximum	50	46	49
Lower 95%	21.2	19.9	20.6
Upper 95%	22.9	21.2	21.9
Number missing	4.0	1.0	7.0

A Bonferroni multiple comparison of means revealed a significant difference only between the Amberley and Richmond groups (p=0.02), with no evidence of difference detected between Amberley and the exposed (p=0.38) or between Richmond and the exposed (p=0.58). This lack of difference between the exposed and comparison groups, especially versus Amberley, seemed to indicate, somewhat paradoxically, that the NART was not as sensitive a measure of education as simply asking the education level in the current study. Also, in a number of analyses for the neuropsychological outcomes, education consistently explained more of the variance than the NART, hence the NART was dropped as a potential confounder.

8.3.6 Depression and anxiety

As part of the General Health and Medical Study, depression and anxiety were assessed as outcomes using the Composite International Diagnostic Interview or CIDI. The CIDI also recorded age of first onset and date of most recent episode, enabling estimation of the prevalence of these for the groups within the last month (see Table 8.14). It is in this latter form that they were used as potential confounders to adjust for mood at the time of the physical examination. There was a strongly significant difference between the groups for

depression within the last month, $\chi^2_2 = 18.94$, $p < 0.0001$, with 12% of the exposed group being classified as depressed compared to 6% from Amberley and 5% from Richmond. There was also a strongly significant difference between the groups for anxiety within the last month, $\chi^2_2 = 33.73$, $p < 0.0001$, with 19% of the exposed group classified as suffering from anxiety compared to 12% from Amberley and 7% from Richmond.

Table 8.14 : Depression and anxiety in the last month as identified by the Composite International Diagnostic Interview (CIDI)

	Amberley		Richmond		Exposed	
	N	%	N	%	N	%
Totals	**406**	***100***	**516**	***100***	**616**	***100***
CIDI DEPRESSION						
Depression	25	*6.2*	28	*5.4*	73	*12*
No Depression	366	*90*	481	*93*	524	*85*
Unknown	15	*3.7*	7	*1.4*	19	*3.1*
CIDI ANXIETY						
Anxiety	50	*12*	38	*7.4*	116	*19*
No Anxiety	341	*84*	471	*91*	481	*78*
Unknown	15	*3.7*	7	*1.4*	19	*3.1*

8.3.7 Civilian occupational exposures

Here, civilian occupational exposure refers to exposure to substances similar to those used in DSRS such as organic solvents. Occupational exposure experienced by the study participants outside their military service was considered an important potential confounder because it was expected that this would vary among the three study groups and that the exposures could be related to study outcomes. The study design assumed that aircraft maintenance would generally be the military occupation of the exposed and Richmond comparison groups but not of the Amberley comparison group. However, civilian occupational exposures could not be extrapolated from military occupation, hence a full occupation history was requested as part of the Postal Questionnaire. Conversion of participants' self-reported occupation history to exposures of interest is detailed below.

8.3.7.1 Measurement

Civilian exposures can be estimated using a Job Exposure Matrix (JEM). A JEM cross-classifies a list of job titles with a list of chemical agents.[3] The use and validity of JEMs has been extensively discussed, with varying conclusions.[4] However, the JEM is currently the most feasible method for assessing occupational exposures in studies involving self-reported occupational history. Additionally, JEMs assign exposure estimates consistently, irrespective of the disease status of the subject, thus decreasing differential information bias. The main disadvantage of general JEMs is their inability to take into account exposure variability within the job categories. By assigning similar exposure to everyone with the same job title, the JEMs may misclassify exposure for a substantial proportion of the subjects under study, and such non-differential misclassification usually attenuates the risk estimates observed.[5]

Recently, a number of studies have used the Finnish Job Exposure Matrix (FINJEM) created by the Finnish Institute of Occupational Health.[6-9] FINJEM is a three-dimensional matrix with job code on axis one, time period categories containing probability of exposure and level of exposure on axis two, and chemical exposure categories on axis three. It was constructed for exposure assessment in large register-based epidemiological studies. The assessment period is 1960-1997, divided into several sub-periods. Exposure is described by the prevalence of exposure and the level of exposure among the exposed, both estimated mainly on continuous scales.[5] To quantify civilian occupational exposure in the current study, three steps were followed:

1) A detailed civilian occupation history was collected.
2) The history was classified to the Australian Standard Classification Occupation (ASCO) and the Australian New Zealand Standard Classification of Industry (ANZSIC).
3) The occupation codes were then translated to the Finnish Job Exposure Matrix Occupation Codes (FINJEM).

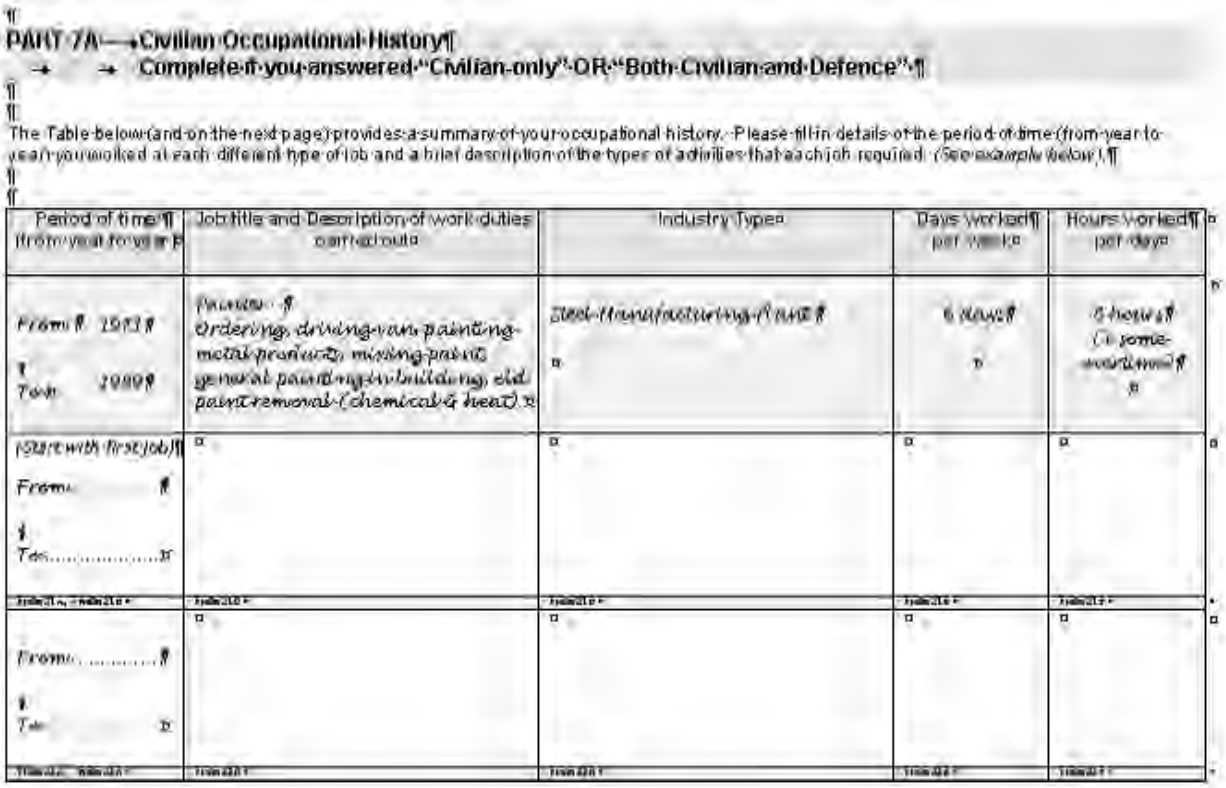
The details are as follows:

[1] A civilian occupational history was collected from respondents in Section 7 of the Postal Questionnaire. Question 7.1 asked respondents to best describe their working life in terms of the following categories:

- civilian only
- Defence Force service only (no civilian jobs held since leaving school)
- both civilian and Defence Force service (in any order of occurrence).

In Section 7A, respondents were then asked to record details of all their civilian employment including start year, finish year, job title and description of work duties carried out, industry type, days worked per week, and hours worked per day. (See Figure 8.1).

Figure 8.1 : Section 7A of Postal Questionnaire



PART 7A — Civilian Occupational History

→ → Complete if you answered "Civilian only" OR "Both Civilian and Defence"

The Table below (and on the next page) provides a summary of your occupational history. Please fill in details of the period of time (from year to year) you worked at each different type of job and a brief description of the types of activities that each job required. (See example below.)

Period of time (from year to year)	Job title and Description of work duties carried out	Industry Types	Days worked per week	Hours worked per day
From 1973 To 1989	Painter Ordering, driving van painting metal products, mixing paint, general painting in building, old paint removal (chemical & heat)	Steel Manufacturing Plant	6 days	8 hours (+ some overtime)
(Start with first job) From ... To ...				
From ... To ...				

[2] The descriptions provided were subsequently classified to the Australian Standard Classification Occupation and the Australian New Zealand Standard Classification of Industry. Occupations were coded to the ASCO using the ASCO Coder[10], a Windows-based structured coding system created by the Australian Bureau of Statistics, which provided a quick and efficient way to code occupation information with a high degree of accuracy and consistency. Industry types were coded to the ANZSIC using the ANZSIC Coder,[11] another Windows-based structured coding system. Like the ASCO coder, the ANZSIC coder provided a quick and efficient way to code industry information to ANZSIC.

[3] The occupation codes were then translated to the Finnish Job Exposure Matrix Occupation Codes (FINJEM). To do this, a member of the study team constructed a concordance tool to convert ASCO codes to FINJEM occupation codes. The FINJEM

provides exposure classification by job title for 39 chemical agents, of which the following ten were determined to be potential confounders for SHOAMP:

Organic solvents	Aliphatic and alicyclic hydrocarbon solvents
	Aromatic hydrocarbon solvents
	Chlorinated hydrocarbon solvents
	Other organic solvents
	Polycyclic Aromatic Hydrocarbons
	Benzo(a)pyrene
Neurotoxins	Lead
Skin sensitisers	Chromium
	Nickel
	Detergent

A study participant was considered exposed to the chemical of interest if they had reported a civilian occupation for which FINJEM reported a probability of exposure greater than 20%. Some study participants reported having no civilian occupations (i.e. only having military service). In these cases the participants were considered to have no civilian exposure.

8.3.7.2 Distribution

A total of 1785 study participants returned a Postal Questionnaire and their responses were subsequently entered into the civilian exposure database and the military history section of the Postal Questionnaire database. Of these participants, 69 were excluded from the analysis as they were considered not exposed to DSRS (see Chapter 4). A further 41 respondents returned Section 7A with insufficient information to code to ASCO and were coded as missing data. The final study population for the analysis of civilian exposure as a potential confounder was 1675.

Table 8.15 describes the proportions in each group with civilian exposures to each category of toxin. From those data, three classifications of civilian occupation exposures were determined for use as potential confounders in analyses:

- any solvent (which included any of the organic solvents and the polycyclic aromatic hydrocarbons)
- lead
- any skin sensitiser.

There were noticeable differences between the groups: 41% of the exposed group reported civilian occupations with solvent exposure, compared to 31% and 17% of the Richmond and Amberley groups respectively, $\chi^2_2 = 70.93, p < 0.0001$. Civilian exposures involving the neurotoxin lead were reported by 44% of the Amberley group, 15% of the Richmond group, and 11% of the exposed group; this difference was significant $\chi^2_2 = 9.53, p = 0.009$. Civilian exposures involving a skin sensitiser were reported by 30% of the Amberley group, 35% of the Richmond group and 43% of the exposed group; this difference was also significant $\chi^2_2 = 20.51, p < 0.0001$.

Table 8.15 : Civilian occupational exposures to potentially confounding toxic substances

Substance	**Amberley**		**Richmond**		**Exposed**	
Total	**N = 473**	***%***	**N = 581**	***%***	**N = 621**	***%***
Organic Solvents						
Alaphatic & Alicyclic Hydrocarbon Solvents	36	*7.6*	24	*4.1*	50	*8.1*
Aromatic Hydrocarbon Solvents	41	*8.7*	40	*6.9*	61	*9.8*
Chlorinated hydrocarbon Solvents	12	*2.5*	6	*1.0*	9	*1.5*
Other Organic Solvents	29	*6.1*	15	*2.6*	31	*5.0*
Polycyclic Aromatic Hydrocarbons (including Benzo(a)pyrene)	24	*5.1*	141	*24*	190	*30*
Any Solvent including PAH	**80**	***17***	**180**	***31***	**252**	***41***
Neurotoxin						
Lead	**44**	***9.3***	**89**	***15***	**70**	***11***
Skin Sensitisers						
Chromium	57	*12*	159	*27*	227	*37*
Nickel	62	*13*	160	*28*	229	*37*
Detergent	97	*21*	163	*28*	217	*35*
Any Skin Sensitiser	**141**	***30***	**203**	***35***	**266**	***43***

8.4 Summary

The preceding descriptive analyses indicate that most common potential confounders are indeed unevenly distributed between the three groups (exposed, Richmond comparison, Amberley comparison). This supports the need to adjust for these variables in all the main analyses where there is a biological rationale for the potential influence of these variables on an outcome of interest.

For the analyses of psychological testing outcomes, education was used rather than the NART for the reasons detailed above.

Other potential confounders specific to a particular outcome – for example, visual acuity as a potential confounder for colour vision – will be dealt with in specific chapters.

8.5 References

1. WHO. WHO Expert Committee on Physical Status. The use and interpretation of anthropometry. (World Health Organisation, Geneva, 1995).
2. Nelson, H.E. & Willison, J.R. *National Adult Reading Test Manual*, (NFER - Nelson, Windsor, UK, 1991).
3. Acheson, E.D. What are job exposure matrices? in *Job Exposure Matrices* (ed. Acheson, E.D.) 1-4 (Medical Research Council, Southhampton, 1982).
4. Acheson, E.D. (ed.) *Job-exposure matrices*, (University of Southhampton, Medical Research Council, Southampton (England), 1983).
5. Kauppinen, T. From cross-tabulations to multipurpose exposure information systems: a new job-exposure matrix. *American Journal of Industrial Medicine* **33**, 409-417 (1998).
6. Alguacil, J. et al. Occupational exposure to dyes, metals, polycyclic aromatic hydrocarbons and other agents and K-ras activation in human exocrine pancreatic cancer. *Int J Cancer* **107**, 635-41 (2003).
7. Alguacil, J. et al. Occupational exposure to organic solvents and K-ras mutations in exocrine pancreatic cancer. *Carcinogenesis* **23**, 101-6 (2002).
8. Weiderpass, E. et al. Occupational exposures and gastrointestinal cancers among Finnish women. *J Occup Environ Med* **45**, 305-15 (2003).
9. Vasama-Neuvonen, K. et al. Ovarian cancer and occupational exposures in Finland. *Am J Ind Med* **36**, 83-9 (1999).
10. Australian Bureau of Statistics. ASCO Coder. Version 4.6.2 edn (Australian Bureau of Statistics, Canberra, 1997).
11. Australian Bureau of Statistics. ANZSIC Coder. Version 2.0 edn (Australian Bureau of Statistics, Canberra, 1997).

9 General Health and Well-Being

Chapter summary

For the study's assessment of general health and well-being, participants were asked to complete a list of 80 self-reported health symptoms, to report any hospitalisation in the past year, and to complete the SF-36 quality of life survey. Additionally, participants in the health examination had a full blood pathology test and urinalysis to check for the presence of protein, glucose and blood. On average, the exposed group self-nominated nearly twice the number of poor health symptoms as the comparison groups. "Feeling unrefreshed after sleep" was a common complaint among all groups. "Forgetfulness" was the most common complaint for the exposed group. Overall, members of the Richmond comparison group were slightly less likely than those of the exposed group to report a hospital admission in the past year. The exposed group recorded poorer quality of life than both comparison groups on both the mental and physical component scores of the SF-36 survey. The blood pathology and urinalysis results were unremarkable.

Chapter contents

9.1 Introduction

Health-related quality of life has been assessed in a variety of settings with military personnel, particularly in response to complaints of adverse health following deployment.[1-4] Rates of hospitalisation were used by Blood and Aboumrad[5] to compare post-conflict health needs of war veterans, as well as to identify the differences in reasons for admission between veterans and non-veterans. Similarly, Knoke and Gray[6] used hospital admission records to identify and compare "unexplained illnesses" for personnel deployed in the Persian Gulf War and those not deployed. Particularly when studying the health of Defence Force personnel, the examination of hospital admission records from different sources (i.e. Veterans' Affairs, Defence and community hospital facilities) can provide a more comprehensive picture of health than military records in isolation.[7]

General health has also been assessed using symptom questionnaires. Self-reported symptoms can be a useful adjunct to the physical examination and they tap into an additional domain of health; symptom questionnaires have been commonly employed during studies of the health of returned servicemen and women from active duty.[3,8-11] The most common complaints from Gray's study of Gulf War "Seebees" were short-term memory problems (39%), unusual fatigue (39%), trouble sleeping (39%), chronic worry/anxiety (38%), and joint stiffness (30%). In addition, the Gulf War group also reported a greater number of hospitalisations and lost workdays compared with the other two groups surveyed. Similarly-developed symptom checklists have been used by Ismail et al.,[12] and Kroenke et al.,[13] during investigations of the health of Gulf War veterans, and by Cherry et al.,[14] Pierce[15] and Wolfe et al.,[16] to document the pattern and extent of ill health of veterans.

In the current study, overall general health and well-being was gauged using four instruments:

1) a list of 80 symptoms experienced by the respondent over the previous month
2) hospital admissions (as indicated by a single yes/no item in the Postal Questionnaire)
3) the Medical Outcomes Study Short Form-36 (SF-36)[17] quality of life survey, which yields quality of life information
4) blood pathology and urinalysis.

9.2 Measures

Symptom studies used to inform the Postal Questionnaire included The Australian Gulf War Veterans' Health Study by Monash University,[18] Derogatis et al.,[19] Ismail et al.,[12] Unwin et al.,[11] and Gray et al.[20]

9.2.1 Self-reported symptoms

The Postal Questionnaire sent to all exposed and comparison individuals as part of their initial General Health and Medical Study (the Health Study) invitation mail-out, contained a list of 80 symptoms indicative of poor health in the past month, to which respondents answered "yes" or "no". A full list of the symptoms is provided in Table 9.1.

Table 9.1 : SHOAMP Postal Questionnaire self-reported symptom checklist

In the past month have you suffered from….		**NO**	**YES**
2.1	Chest pain	o	o
2.2	Headaches	o	o
2.3	Rapid, pounding or irregular heart beat	o	o
2.4	Irritability / outbursts of anger	o	o
2.5	Shortness of breath	o	o
2.6	Wheezing	o	o
2.7	Sleeping difficulties	o	o
2.8	Feeling jumpy / easily startled	o	o
2.9	Feeling unrefreshed after sleep	o	o
2.10	Fatigue	o	o
2.11	Double vision	o	o
2.12	Itchy or painful eyes	o	o
2.13	Rash or skin irritation	o	o
2.14	Skin ulcers	o	o
2.15	Feeling distant or cut off from others	o	o
2.16	Constipation	o	o
2.17	Flatulence or burping	o	o
2.18	Stomach cramps	o	o
2.19	Diarrhoea	o	o
2.20	Indigestion	o	o
2.21	Dry mouth	o	o

Table 9.1 continued…

In the past month have you suffered from….		NO	YES
2.22	Mouth ulcers	o	o
2.23	Toothache	o	o
2.24	Persistent cough	o	o
2.25	Any new lump(s) in the breast area	o	o
2.26	Any change to the skin of nipple/breast	o	o
2.27	An unusual increase in the size of one breast	o	o
2.28	Sticky or bloody discharge from one/both nipples	o	o
2.29	Sore throat	o	o
2.30	Forgetfulness	o	o
2.31	Dizziness	o	o
2.32	Seizures or convulsions	o	o
2.33	Fainting	o	o
2.34	Blackouts	o	o
2.35	Feeling disorientated	o	o
2.36	Loss of concentration	o	o
2.37	Difficulty finding the right word	o	o
2.38	Pain on passing urine	o	o
2.39	Passing urine more often	o	o
2.40	Loss of control over bladder or bowels	o	o
2.41	Bleeding during bowel movements	o	o
2.42	Loss of interest in sex	o	o
2.43	Problems with sexual functioning	o	o
2.44	Increased sensitivity to light	o	o
2.45	Increased sensitivity to noise	o	o
2.46	Increased sensitivity to smells or odours	o	o
2.47	Ringing in the ears	o	o
2.48	Avoiding doing things or situations	o	o
2.49	Pain, without swelling or redness, in several joints	o	o
2.50	Stiffness in several joints	o	o
2.51	General muscle aches or pains	o	o
2.52	Loss of balance or coordination	o	o
2.53	Difficulty speaking	o	o
2.54	Low back pain	o	o
2.55	Night sweats which soak the bed sheets	o	o
2.56	Feeling feverish	o	o
2.57	Tender or painful swelling of lymph glands in neck, armpit or groin	o	o
2.58	Loss of, or decrease in, appetite	o	o

Table 9.1 continued…

In the past month have you suffered from….		NO	YES
2.59	Nausea	o	o
2.60	Vomiting	o	o
2.61	Distressing dreams	o	o
2.62	Unintended weight gain greater than 4 kg	o	o
2.63	Unintended weight loss greater than 4 kg	o	o
2.64	Difficulty lifting objects above your head, or from a high shelf	o	o
2.65	Difficulty undoing buttons	o	o
2.66	Difficulty turning doorknobs or unscrewing jars	o	o
2.67	Difficulty getting up from sitting in a chair or couch	o	o
2.68	Problems with tripping, or your feet slapping, while walking	o	o
2.69	Difficulty recognising hot from cold water	o	o
2.70	Difficulty feeling pain, cuts or injuries	o	o
2.71	Feeling unsteady walking on uneven ground	o	o
2.72	Feeling unsteady walking in the dark	o	o
2.73	Feeling like you may fall over because of your unsteadiness	o	o
2.74	Numbness, “asleep feeling” or prickling sensation in hands or arms	o	o
2.75	Numbness, “asleep feeling” or prickling sensation in feet or legs	o	o
2.76	Burning, deep aching pain, or tenderness in your hands or arms	o	o
2.77	Burning, deep aching pain, or tenderness in your feet or legs	o	o
2.78	Unusual sensitivity or tenderness of your skin when clothes or bedclothes rub against you	o	o
2.79	Feeling like you will faint, or fainting, when you stand up from a lying or sitting position	o	o
2.80	Difficulty swallowing food (more than occasionally)	o	o

9.2.2 Hospitalisation

The Postal Questionnaire contained a single item which asked participants to report whether or not they had “been to hospital for an overnight stay or longer” during the past 12 months. If they had been admitted to hospital, they had then to make a note of how many times. No specific information regarding hospitalisation was collected during the health examination, unless a participant wished to discuss their health situation with Health Services Australia (HAS) clinicians or nursing staff and have it noted in their Health Examination Booklet.

9.2.3 Quality of life

The SF-36 quality of life survey is a standardised multi-dimensional measure of self-perceived general health status that has been validated in adult populations in the United States, Great Britain, and Australia.[17,21-23] The scale measures eight health-related concepts, and two summary component scores are also compiled which represent mental and physical well-being. High reliability has been demonstrated for all sub-scales (Cronbach's alpha>0.80), and factor analysis confirmed construct and criterion validity.[24,25] The survey was constructed for self-administration and is also suitable for administration by a trained interviewer in person or by telephone; there are some differences in results obtained from the two methods.[22,26,27]

The Postal Questionnaire sent out to exposed and comparison individuals contained the SF-36 quality of life survey, specially printed in a different colour to make it stand out. It was hoped that this quality of life data would be returned not only by participants who chose to complete the entire questionnaire and/or take part in the health examination, but also by those who refused to do so, in order that they could at least provide an indication of their overall well-being.

The SF-36 was designed to be used in clinical practice and research, health policy evaluations, and general population surveys. The SF-36 assesses eight health concepts:

- limitations in physical activities because of health problems
- limitations in social activities because of physical or emotional problems
- limitations in usual role activities because of physical health problems
- bodily pain
- general mental health (psychological distress and well-being)
- limitations in usual role activities because of emotional problems
- vitality (energy and fatigue)
- general health perceptions.

These are scored from 0 to 100, with a higher score indicating better self-reported health. These scales are combined to yield two summary component scores, one for physical and one for mental well-being.

9.2.4 Blood pathology and urine test results

As previously described in Chapter 5 (Measures), blood was collected from consenting participants during the health examination according to instructions provided in the Nurse Instruction Manual (Appendix 7A). Tests included full blood examination, liver function tests, electrolytes and urea, calcium and phosphate, random glucose and cholesterol, C-reactive protein, Apoliproprotein E and Homocysteine. Urinanalysis was also conducted for the detection of glucose (at 30 seconds) and protein and blood (at 60 seconds).

9.3 Potential confounders

Potential confounders for analyses of hospital admission in the past year were age, posting category, rank, smoking status, alcohol intake, and BMI category. Cigarette smoking, alcohol consumption and high BMI have all been shown to be associated with a range of adverse health outcomes and hospital admissions.

Potential confounders for SF-36 analyses were age, posting category, rank category, smoking status, alcohol intake, education, and civilian solvent exposure.

Potential confounders considered for blood pathology were age, posting date, rank, alcohol behaviour, smoking behaviour, civilian solvent exposure, and HSA centre.

Confounders were not considered relevant for self-reported symptoms as no regression analyses were conducted on these outcomes.

9.4 Analyses

All analyses followed the outline described in Chapter 6 (Analysis), except for the self-reported symptoms and blood pathology. A descriptive analysis was undertaken for the 80 Postal Questionnaire items of self-reported symptoms. Firstly, a chi-square value (see Glossary) was obtained to compare the presence/absence of symptoms across groups. Items were then ranked according to the chi-square value and inspected to confirm that the proportion with self-reported symptoms was higher for the exposed than for the Richmond

and Amberley groups. The top ten self-reported items were selected for each of the three groups by frequency. These were qualitatively compared with each other and with the top ten ranked by chi-square.

The proportion of individuals in each exposure group reporting that they had been to hospital for an overnight stay or longer during the past 12 months was presented, and logistic regression was used for primary and secondary analysis as described in Chapter 6 (Analysis).

The SF-36 profile scores were calculated according to the method described in Ware [28] and were plotted according to the method described in the National Health Survey 1997 [23]; the latter involves comparing results to the general Australian profiles as a reference. Component summary scores for physical health and mental health were calculated by computing Z-scores using Australian age-group means and standard deviations.[29] The standardised scores were summed, with each profile score weighted by the factor analytic scores from the National Health Survey 1997[23] according to the method described therein. These were then standardised to a mean of 50 and a standard deviation of 10, allowing comparison with the Australian population. The physical component score and the mental component score were then analysed using linear regression as outlined in Chapter 6 (Analysis).

Blood pathology test results were analysed in two ways. Means were compared in a general linear model including potential confounders. Additionally, the number and proportion with values less/greater than normal and within normal ranges (according to defined laboratory definitions)[30,31] were identified, and these were analysed in a 3 x 3 contingency table using Fisher's exact test, as many cell frequencies were small. Urinalysis was analysed in a 2 x 3 contingency table using Fisher's exact test. For this analysis, negative result and trace were combined; for urinalysis for protein and glucose, the categories 1+, 2+, 3+ and 4+ were combined; while for urinalysis for blood, the categories small, moderate and large were combined.

9.5 Results

9.5.1 Self-reported symptoms

There were between 15 and 28 missing responses to each of the 80 questions relating to self-report of symptoms. Among respondents, there were 1623 who answered all questions, 52 who completed all but one question, 39 who answered some or most questions, and 12 who did not answer any of the questions. Of the questions answered, an average of about 23 symptoms were identified by the exposed group compared to about 14 by the Richmond group and about 15 by the Amberley (Table 9.2). The maximum number of symptoms reported was 68 from an exposed participant compared to 64 from a Richmond participant and 59 from an Amberley participant.

Table 9.2 : Total number of self-reported symptoms – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	14.6	13.9	23.1
Standard Deviation	11.6	10.6	14.6
50th Percentile	12.0	12.0	21.0
Minimum	0.0	0.0	0.0
Maximum	59.0	64.0	68.0
Lower 95% CL for Mean	13.6	13.1	21.9
Upper 95% CL for Mean	15.7	14.8	24.3
Number missing	1.0	1.0	0.0

The top ten self-report items that differed most between the exposed and comparison groups (as measured by chi-square values) are presented in Table 9.3. From this table it can be seen that the top three self-report items were “loss of concentration”, “forgetfulness” and “difficulty finding the right word”. For all top ten items, the proportion of the exposed group that reported in the affirmative was much higher than that of both the Amberley and Richmond comparisons. Table 9.4, Table 9.5 and Table 9.6 detail the top ten most common symptoms by frequency for the Amberley, Richmond and exposed groups respectively (complete lists are located in Appendix 9). “Fatigue” and “feeling unrefreshed after sleep” were the top two items for Richmond and Amberley, while “forgetfulness” and “feeling unrefreshed after sleep” were the top two for the exposed group.

Table 9.3 : Top 10 self-reported symptoms, ranked by chi square

Item	Amberley N	Amberley %	Richmond N	Richmond %	Exposed N	Exposed %	Total	No. Missing	Chi Square
Loss of concentration	164	34	221	38	438	69	1710	6	178.83
Forgetfulness	198	41	259	44	472	74	1711	5	164.51
Difficulty finding the right word	203	42	267	45	448	71	1711	5	117.38
Irritability / outbursts of anger	224	46	255	43	436	69	1708	8	95.58
Feeling disorientated	37	7.6	41	7.0	151	24	1707	9	94.05
Feeling jumpy / easily startled	106	22	106	18	258	41	1709	7	89.93
Feeling distant or cut off from others	102	21	108	18	250	39	1709	7	81.78
Dizziness	73	15	93	16	207	33	1709	7	68.81
Rash or skin irritation	145	30	196	33	328	52	1709	7	68.38
Feeling unsteady walking in the dark	41	8.5	50	8.5	142	22	1707	9	66.22

Table 9.4 : Top 10 self-reported symptoms, for Amberley group ranked by frequency

Item	Amberley N	Amberley %	Richmond N	Richmond %	Exposed N	Exposed %	Total	No. Missing	Chi Square
Fatigue	**294**	**60**	324	55	469	74	1709	7	51.08
Feeling unrefreshed after sleep	**293**	**60**	332	56	471	74	1711	5	47.38
Headaches	**285**	**59**	318	54	454	71	1711	5	42.71
Low back pain	**258**	**53**	323	55	408	64	1711	5	18.54
Sleeping difficulties	**257**	**53**	300	51	437	69	1711	5	48.00
General muscle aches or pains	**250**	**51**	311	53	412	65	1710	6	27.62
Irritability / outbursts of anger	**224**	**46**	255	43	436	69	1708	8	95.58
Flatulence or burping	**223**	**46**	261	44	390	62	1710	6	43.87
Stiffness in several joints	**209**	**43**	246	42	352	56	1711	5	28.92
Difficulty finding the right word	**203**	**42**	267	45	448	71	1711	5	117.38

Table 9.5 : Top 10 self-reported symptoms, for Richmond group ranked by frequency

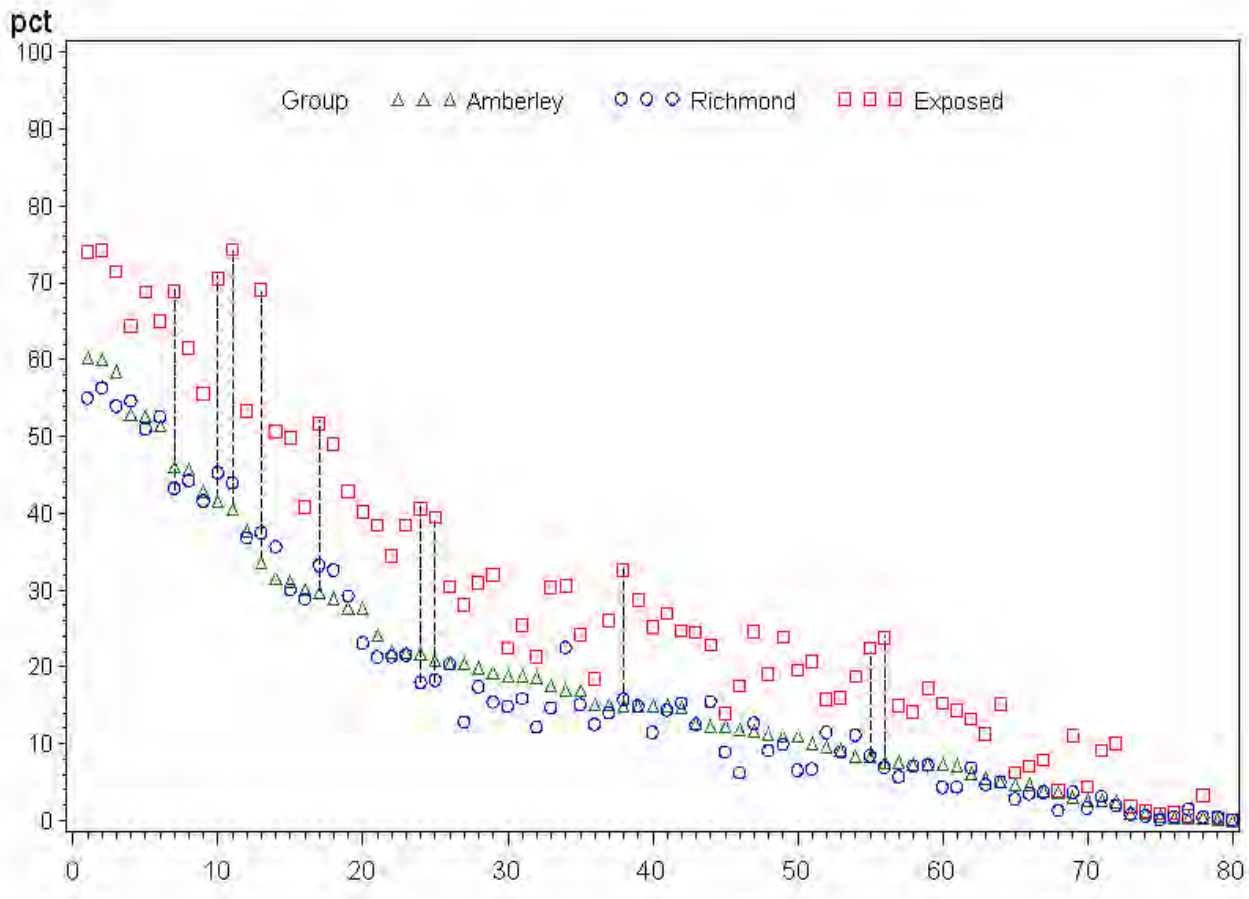
Item	Amberley N	Amberley %	Richmond N	Richmond %	Exposed N	Exposed %	Total	No. Missing	Chi Square
Feeling unrefreshed after sleep	293	60	**332**	**56**	471	74	1711	5	47.38
Fatigue	294	60	**324**	**55**	469	74	1709	7	51.08
Low back pain	258	53	**323**	**55**	408	64	1711	5	18.54
Headaches	285	59	**318**	**54**	454	71	1711	5	42.71
General muscle aches or pains	250	51	**311**	**53**	412	65	1710	6	27.62
Sleeping difficulties	257	53	**300**	**51**	437	69	1711	5	48.00
Difficulty finding the right word	203	42	**267**	**45**	448	71	1711	5	117.38
Flatulence or burping	223	46	**261**	**44**	390	62	1710	6	43.87
Forgetfulness	198	41	**259**	**44**	472	74	1711	5	164.51
Irritability / outbursts of anger	224	46	**255**	**43**	436	69	1708	8	95.58

Table 9.6 : Top ten self-reported symptoms, for exposed group ranked by frequency

Item	Amberley N	Amberley %	Richmond N	Richmond %	Exposed N	Exposed %	Total	No. Missing	Chi Square
Forgetfulness	198	41	259	44	**472**	**74**	1711	5	164.51
Feeling unrefreshed after sleep	293	60	332	56	**471**	**74**	1711	5	47.38
Fatigue	294	60	324	55	**469**	**74**	1709	7	51.08
Headaches	285	59	318	54	**454**	**71**	1711	5	42.71
Difficulty finding the right word	203	42	267	45	**448**	**71**	1711	5	117.38
Loss of concentration	164	34	221	38	**438**	**69**	1710	6	178.83
Irritability / outbursts of anger	224	46	255	43	**436**	**69**	1708	8	95.58
Sleeping difficulties	257	53	300	51	**437**	**69**	1711	5	48.00
General muscle aches or pains	250	51	311	53	**412**	**65**	1710	6	27.62
Low back pain	258	53	323	55	**408**	**64**	1711	5	18.54

Figure 9.1 plots the frequency of all 80 symptoms for all three groups, ordered by decreasing frequency of symptoms in the Amberley group. It shows that the exposed group generally reports a higher frequency for almost all symptoms. One would expect that if the exposed group were simply over-reporting symptoms non-specifically, the line along which the points for the exposed group lie would be monotonic from left to right (i.e. unvarying from left to right). However, there are a number of symptoms for which the frequency is specifically high, i.e. the points do not lie on the same line as the rest of the symptoms. This argues for a specific effect in the F-111 DSRS group and not just a "general" increase in complaints reported.

Figure 9.1 : Affirmative responses to the Postal Questionnaire self-reported symptom items, ranked by Amberley frequency



Vertical lines = top 10 chi-squared differences between exposed (upper line with squares) and the comparison groups (lower triangles and circles).

9.5.2 Hospitalisation

Five people did not complete the question on hospital admissions in the past 12 months. Table 9.7 shows the proportion of individuals in each group reporting a hospital admission in the past 12 months. The proportion of those in the exposed group was similar to that of the Amberley comparison group, with a slightly lower percentage in the Richmond comparison group.

Table 9.7 : Number and percentage of participants reporting a hospital admission within the past 12 months by exposure group

Hospital admission	Amberley N = 485		Richmond N = 591		Exposed N = 635	
	n	%	n	%	n	%
Any hospital admission	80	*16*	74	*13*	103	*16*

Regression analyses were conducted on the 1492 individuals who had complete data (individuals who did not take part in the health examination had missing data for BMI). In the primary analysis, overall exposure group was not statistically significantly associated with self-reported hospital admission within the past 12 months (p = 0.096) (see Table 9.8). However, the point estimate of the odds ratio for the Richmond comparison group relative to the exposed group was significant at 0.68 with 95% CI (0.48, 0.97). This indicates a slightly lower odds of admission for the Richmond comparison group compared to the exposed group. There was no significant association between exposure group and hospital admission for Program 1, although the Richmond comparison came close, with the 95% confidence interval for that odds ratio just including one. This association was significant for the Program 2 subgroup analysis, with an odds ratio of 0.55 and 95% CI (0.35, 0.89). There was little evidence of a dose-response relationship, with odds ratios of 0.74, 1.51 and 1.39 for mild, moderate and prolonged exposure respectively, relative to unexposed, and this was not significant (p=0.068).

Table 9.8 : Any self-reported hospital admission within the past 12 months – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

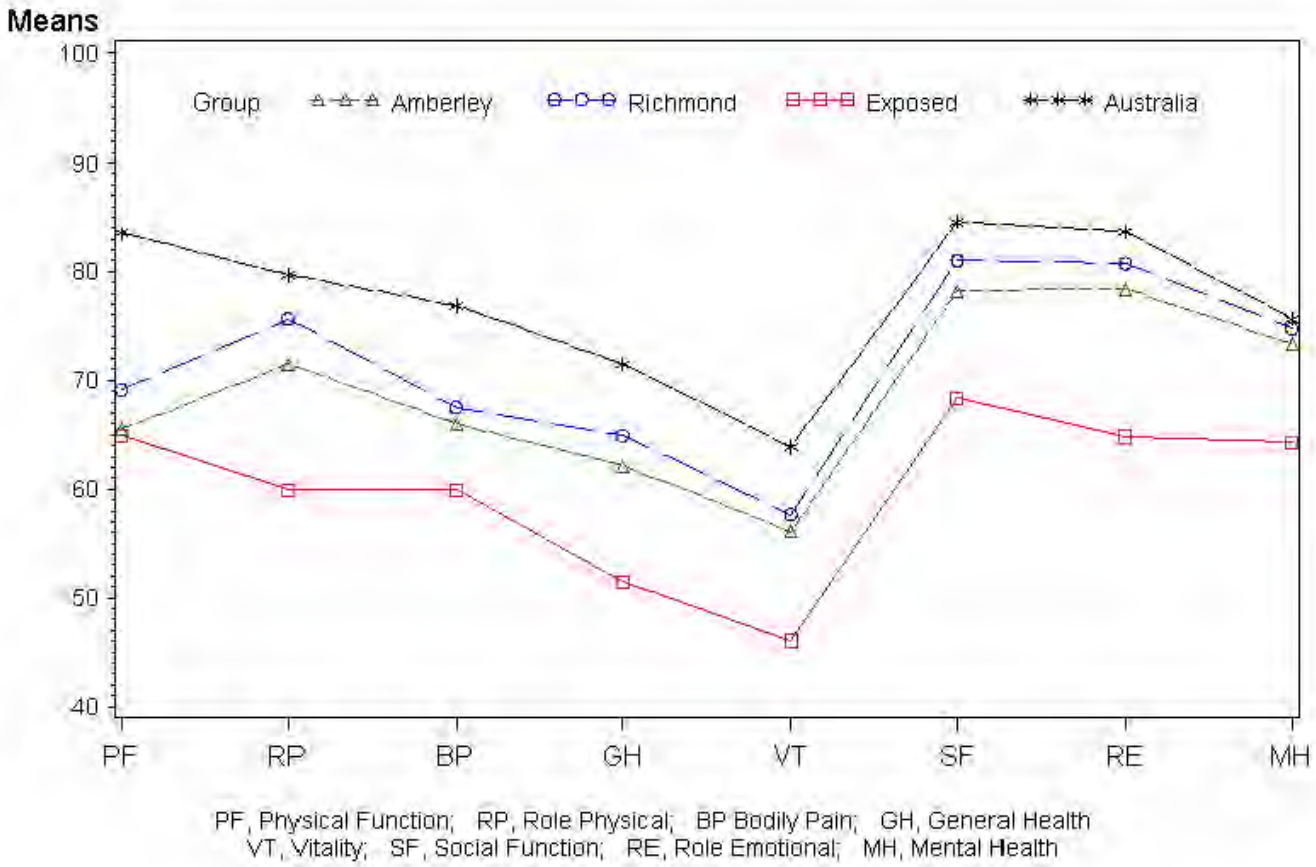
Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.93	0.65	1.33	2	4.69	0.0956
	Richmond vs Exposed	0.68	0.48	0.97	.	.	.
Program 1	Amberley vs Exposed	0.91	0.58	1.42	2	3.77	0.1518
	Richmond vs Exposed	0.67	0.43	1.05	.	.	.
Program 2	Amberley vs Exposed	0.77	0.48	1.22	2	6.48	0.0392
	Richmond vs Exposed	0.55	0.35	0.89	.	.	.
Dose	Mild exposure vs Unexposed	0.74	0.43	1.28	3	7.12	0.0682
	Moderate exposure vs Unexposed	1.51	0.98	2.35	.	.	.
	Prolonged exposure vs Unexposed	1.39	0.89	2.15	.	.	.

"Mild exposure" = up to 9 months. "Moderate exposure" = 10 to 29 months. "Prolonged exposure" = 30 months or more (see Chapter 4).

9.5.3 Quality of life

It is customary to graphically present the eight quality of life dimensions of health in a line plot with the dimensions ordered from physical health to mental health. The points are joined by a line to facilitate comparisons between profiles; however, the dimensions are independent. Comparisons can be made within dimensions but not between.[23] Presented in Figure 9.2 are the ordered health dimensions.

Figure 9.2 : Comparison of SHOAMP group SF-36 profiles against Australian norms



Here it can be seen that the exposed group is lower than both Amberley and Richmond on all dimensions, except physical functioning (PF) where the three groups appear quite similar. Interestingly, the exposed group (lowest line, squares) is parallel to, and lower than, the Australian norms (uppermost line, stars) for all dimensions; the two comparison groups are also parallel to the Australian norms except for physical functioning. On the whole, the two comparison groups are noticeably lower than the Australian norms for all of the physical functioning dimensions, but get closer for the mental health dimensions, and are closest at the overall mental health dimension.

The mean physical component score for the exposed group was lower than the means of the two comparison groups: 41 compared to 42 for Amberley and 44 for Richmond (Table 9.9). There was a significant group association when considering all potential confounders in multiple linear regression for all exposed (p=0.004), Program 1 (p=0.0007) and Program 2 (p=0.040) (see Table 9.10). For the primary analysis Amberley scored on average 1.23 (-0.31, 2.77) points higher than the exposed, which was not significant, and Richmond scored on average 2.27 (0.95, 3.60) points higher than the exposed, which was significant.

Table 9.9 : Physical Component Score – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	41.9	43.8	41.4
Standard Deviation	11.9	11.0	11.2
50th Percentile	44.0	45.8	41.6
Lower Quartile	32.6	35.7	33.2
Upper Quartile	51.9	52.5	50.5

Table 9.10 : Physical Component Score – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	1.23	-0.31	2.77	0.05	2	0.0035
	Richmond	2.27	0.95	3.60			
Program 1	Amberley	2.40	0.50	4.30	0.06	2	0.0007
	Richmond	3.25	1.58	4.92			
Program 2	Amberley	1.58	-0.43	3.60	0.06	2	0.0395
	Richmond	2.36	0.54	4.18			
Dose	Mild exposure	-0.47	-2.38	1.43	0.05	3	0.0201
	Moderate exposure	-1.89	-3.72	-0.05			
	Prolonged exposure	-2.48	-4.25	-0.71			

A stronger association was observed for Program 1, where Amberley scored on average 2.40 (0.50, 4.30) points higher than the exposed, and Richmond scored on average 3.25 (1.58, 4.92) points higher. The association was quite weak for Program 2, where the Amberley group scored on average 1.58 (-0.43, 3.60) points higher than the exposed, which was not significant, and Richmond scored 2.36 (0.54, 4.18) points higher than the exposed, which was significant. The 95% confidence interval for Amberley spanned zero for all exposed and Program 2. There was an association between dose and the physical component score (p=0.020): participants classified in the mild exposure group scored on average 0.47 (-2.38, 1.43) points lower than the unexposed; the moderate exposure group

scored on average 1.89 (-3.72, -0.05) points lower than the unexposed; and the prolonged exposure group scored on average 2.48 (-4.25, -0.71). The latter two results were significant, with the 95% confidence intervals not including zero.

The mean mental health component score for the exposed group was lower than for the two comparison groups: 43 compared to 49 for Amberley and 50 for Richmond (Table 9.11).

Table 9.11 : Mental Component Score – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	49.4	49.9	42.9
Standard Deviation	13.4	12.3	14.7
50th Percentile	53.1	53.5	46.2
Lower Quartile	42.9	43.8	31.6
Upper Quartile	58.9	58.2	54.7

There was a significant group association when considering all potential confounders in multiple linear regression for all exposed ($p<0.0001$), Program 1 ($p<0.0001$) and Program 2 ($p<0.0001$) (see Table 9.12). Specifically, for the primary analysis, Amberley scored on average 7.22 (5.40, 9.04) points higher than the exposed, and Richmond scored on average 6.50 (4.93, 8.06) points higher than the exposed. A similar strong association was observed for Program 1, where Amberley scored on average 7.67 (5.44, 9.90) points higher than the exposed, and Richmond scored on average 7.25 (5.30, 9.21) points higher than the exposed. The association also remained strong for Program 2, where the Amberley group scored on average 6.39 (4.09, 8.69) points higher than the exposed, and Richmond scored 5.70 (3.62, 7.79) points higher than the exposed.

Table 9.12 : Mental Component Score – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	7.22	5.40	9.04	0.10	2	<.0001
	Richmond	6.50	4.93	8.06			
Program 1	Amberley	7.67	5.44	9.90	0.11	2	<.0001
	Richmond	7.25	5.30	9.21			
Program 2	Amberley	6.39	4.09	8.69	0.08	2	<.0001
	Richmond	5.70	3.62	7.79			
Dose	Mild exposure	-4.17	-6.40	-1.95	0.10	3	<.0001
	Moderate exposure	-7.92	-10.07	-5.78			
	Prolonged exposure	-7.59	-9.66	-5.52			

There was an association between dose and the mental component score ($p<0.0001$), although a monotonic dose-response relationship was not observed: participants classified in the mild exposure group scored on average 4.17 (-6.40, -1.95) points lower than the unexposed; the moderate exposure group scored on average 7.92 (-10.07, -5.78) points lower than the unexposed; and the prolonged exposure group scored on average 7.59 (-9.66, -5.52) points lower.

9.5.4 Blood pathology and urine test results

For none of the pathology tests were there any significant differences detected between the exposed group and either of the comparison groups by the general linear model or by the contingency table analysis. Where a difference was detected, this difference was either between the Richmond and Amberley groups or due to these groups showing an abnormal result (e.g. no out-of-range values for the exposed compared to several out-of-range values for the two comparison groups). See Appendix 9T for blood rest results and normal ranges.[30,31] See Appendices 9U, 9V and 9W for urine test results.

9.6 Discussion

General well-being was measured using a number of methods:

1) Firstly, a simple inventory of 80 symptoms. These were judged qualitatively and quantitatively. It is apparent that the list of the top ten symptoms for the Amberley and Richmond comparison groups contains the vague and non-specific symptoms that one would expect are common in the general population; indeed, these are also documented in previous military studies.[10,14,32] They include fatigue, feeling unrefreshed after sleep, headaches, general muscle aches and pains, sleeping difficulties, and flatulence or burping. In contrast, the top ten ***differences*** between the list of symptoms in the exposed group compared with the comparison groups include such specific complaints as loss of concentration; forgetfulness; difficulty finding the right word; feeling disoriented, jumpy, or cut off from others; skin rash; and feeling unsteady walking in the dark. The specificity of these symptoms supports the interpretation that they are not simply an over-reporting of common complaints but are "unique" to the exposed group.

2) Secondly, hospitalisations. The exposed group appears to have a statistically significant increase in self-reported hospitalisations compared to the Richmond group (1.47, 95% CI 1.03-2.08) but not compared to the Amberley group. This result is consistent in subgroup analyses of Program 1 and 2, but there is no clear dose-response curve. It is important to note that due to time constraints we did not ask the indications for hospitalisation, nor the length of stay. Hence we are simply using this as an indication of quality of life or disease "burden" in general. It is possible that between-group differences in hospitalisations reflect between-State variability in hospital bed occupancy and admission practices.

3) Thirdly, the well-validated and widely-used SF-36 scale. This scale has eight sub-domains which form two summary component scores: a physical component score and a mental component score. These two component scores are adjusted to the Australian population so that the normative score is 50, with a standard deviation of 10. On the physical scores, all three groups score worse than the Australian population, with scores between 41 and 44 compared to 50. This would place the Richmond group in the 30th percentile for physical health, and the Amberley and exposed groups in the 20th percentile. These scores are consistent and statistically different between the

exposed and Richmond group overall and in both subgroup analyses, and follow a dose-response curve. On the mental scores, the Amberley and Richmond groups score almost exactly at the 50th percentile, whereas the exposed group scores seven points lower, placing them in the 30th percentile. These between-group differences are consistent and statistically significant for the overall comparison and for both subgroup comparisons, and there is the suggestion of a dose-response curve.

4) Fourthly, screening blood tests and urinalysis. These tests covered electrolytes, full blood count, kidney function, liver function, cholesterol, and a marker of inflammation, and there were no differences between the exposed group and either of the comparisons on any of these tests.

9.7 Conclusions

In summary, taken as a whole these results indicate a significantly lower quality of life for the exposed group. Those who participated in DSRS activities report not only a greater number of symptoms but also a predominance of mental health symptoms. This is moderately consistent with an increased number of hospitalisations, but more importantly it is strongly consistent with the results from the SF-36, indicating poorer quality of life in the physical domain and to a greater extent in the mental/emotional domain. This translates into a one or two decile drop in quality of life, placing the exposed group on average in the bottom 20-30% of the Australian population for quality of life.

9.8 References

1. Black, D.W. et al. Quality of life and health-services utilisation in a population-based sample of military personnel reporting multiple chemical sensitivities. *Journal of Occupational and Environmental Medicine* **41**, 928-933 (1999).
2. Voelker, M.D. et al. Health-related quality of life in Gulf War era military personnel. *American Journal of Epidemiology* **155**, 899-907 (2002).
3. The Iowa Persian Gulf Study Group. Self-reported illness and health status among Persian Gulf War veterans: a population-based study. *Journal of the American Medical Association* **277**, 238-245 (1997).
4. Barrett, D.H., Boehmer, T.K., Boothe, V.L., Flanders, W.D. & Barrett, D.H. Health-related quality of life of U.S. military personnel: a population-based study. *Military Medicine* **168**, 941-947 (2003).
5. Blood, C.G. & Aboumrad, T.L. A comparison of postdeployment hospitalization incidence between active duty Vietnam and Persian Gulf War veterans. *Military Medicine* **166**, 648-655 (2001).
6. Knoke, J.D. & Gray, G.C. Hospitalizations for unexplained illnesses among U.S. veterans of the Persian Gulf War. *Emerging Infectious Diseases* **4**, 211-219 (1998).
7. Gray, G.C., Smith, T.C., Kang, H.K. & Knoke, J.D. Are Gulf War veterans suffering war-related illnesses? Federal and civilian hospitalizations examined June 1991 to December 1994. *American Journal of Epidemiology* **151**, 63-71 (2000).
8. Joseph, S.C. A comprehensive clinical evaluation of 20,000 Persian Gulf War veterans. Comprehensive Clinical Evaluation Program Evaluation Team. *Military Medicine* **162**, 149-155 (1997).
9. Haley, R.W., Kurt, T.L. & Hom, J. Is there a Gulf War Syndrome? Searching for syndromes by factor analysis of symptoms. *Journal of the American Medical Association* **277**, 215-222 (1997).
10. Gray, G.C., Kaiser, K.S., Hawksworth, A.W., Hall, F.W. & Barrett-Connor, E. Increased postwar symptoms and psychological morbidity among U.S. Navy Gulf War veterans. *American Journal of Tropical Medicine and Hygiene* **60**, 758-766 (1999).
11. Unwin, C. et al. Health of UK servicemen who served in Persian Gulf War. *The Lancet* **353**, 169-178 (1999).
12. Ismail, K. et al. Is there a Gulf War syndrome? *The Lancet* **353**, 179-182 (1999).
13. Kroenke, K., Koslowe, P. & Roy, M. Symptoms in 18,495 Persian Gulf War veterans. *Journal of Occupational and Environmental Medicine* **40**, 520-528 (1998).

14. Cherry, N. et al. Health and exposures of United Kingdom Gulf War veterans. Part I: the pattern and extent of ill health. *Occupational and Environmental Medicine* **58**, 291-298 (2001).
15. Pierce, P.F. Physical and emotional health of Gulf War veteran women. *Aviation, Space and Environmental Medicine* **68**, 317-321 (1997).
16. Wolfe, J., Proctor, S.P., Davis, J.D., Borgos, M.S. & Friedman, M.J. Health symptoms reported by Persian Gulf War veterans two years after return. *American Journal of Industrial Medicine* **33**, 104-113 (1998).
17. Ware, J.E. & Sherbourne, C.D. The MOS 36-item Short Form Health Survey (SF-36): I. Conceptual framework and item selection. *Medical Care* **30**, 473-483 (1992).
18. Monash University. Australian Gulf War Veterans' Health Study. Vol III (Monash University, Melbourne, Vic, 2003).
19. Derogatis, L., Lipman, R., Rickels, K., Uhlenhuth, E. & Covi, L. The Hopkins Symptom Checklist: A self report symptom inventory. *Behavioural Science* **19**, 1-15 (1974).
20. Gray, G.C., Reed, R.J., Kaiser, K.S., Smith, T.C. & Gastanaga, V.M. Self-reported symptoms and medical conditions among 11,868 Gulf War-era veterans. *American Journal of Epidemiology* **155**, 1033-1044 (2002).
21. McCallum, J. The SF-36 in an Australian sample: Validating a new, generic health status measure. *Australian Journal of Public Health* **19**, 160-166 (1995).
22. Watson, E.K., Firman, D.W., Baade, P.D. & Ring, I. Telephone administration of the SF-36 health survey: Validation studies and population norms for adults in Queensland. *Australian and New Zealand Journal of Public Health* **20**, 359-363 (1996).
23. Australian Bureau of Statistics. National Health Survey. SF-36 Population Norms. 1995. (Australian Bureau of Statistics, Canberra, 1997).
24. Garratt, A.M., Ruta, D.A., Abdalla, M.I., Buckingham, J.K. & Russell, I.T. The SF-36 health survey questionnaire: An outcome measure suitable for routine use within the NHS? *British Medical Journal* **306**, 1440-1444 (1993).
25. Jenkinson, C., Wright, L. & Coulter, A. Criterion validity and reliability of the SF-36 in a population sample. *Quality of Life Research* **3**, 7-12 (1994).
26. Perkins, J.J. & Sanson-Fisher, R.W. An examination of self- and telephone-administered modes of administration for the Australian SF-36. *Journal of Clinical Epidemiology* **51**, 969-973 (1998).
27. McHorney, C.A., Kosinski, M. & Ware, J.E.J. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Medical Care* **32**, 551-567 (1994).
28. Ware, J.E., Snow, K.K., Kosinski, M. & Gandek, B. *SF-36 Health Survey. Manual and Interpretation Guide.*, (QualityMetric Incorporated, Lincoln, 2002).

29. Stevenson, C. SF-36: Interim norms for Australian data. 72 (Australian Institute of Health and Welfare, Canberra, 1996).
30. The Royal College of Pathologists of Australasia. *Manual of use and interpretation of pathology tests.*, (Sydney, Australia, 1997).
31. Wallach, J. *Interpretation of diagnostic tests. A synopsis of laboratory medicine.*, (Little, Brown and Company., Boston, USA., 1992).
32. Kang, H.K. et al. Evidence for a deployment-related Gulf War syndrome by factor analysis. *Archives of Environmental Health* **57**, 61-68 (2002).

10 Cardiovascular Health

Chapter summary

The General Medical and Health Study focused on three potential cardiac effects of organic solvents: (a) palpitations, (b) autonomic nervous system, and (c) heart disease. Tested by sitting and standing blood pressure, postural hypotension was detected in approximately 1% of participants only. Self-reported symptoms of dizziness and/or feeling faint were both consistently and significantly different in the exposed group compared to Amberley and Richmond. Both these symptoms were statistically significant for Programs 1 and 2 and showed a dose-response effect. Self-reported chest pain and heart palpitations were roughly twice as common in the exposed group than the Amberley and Richmond groups. Again, both of these were significant in the Program 1 and 2 subgroup comparisons and both showed a dose-response effect. There were no significant differences between the exposed and comparison groups in terms of self-reported physician diagnosis of high blood pressure or heart disease.

Chapter contents

10.1 Introduction

There were three main cardiovascular concerns for SHOAMP. They were:

1) The reported association between exposure to organic solvents and the autonomic nervous system, particularly autonomic neuropathy which in turn can cause postural hypotension.[1,2] The autonomic nervous system is that part of the nervous system which controls such involuntary functions as the regulation of the activity of smooth muscles, the heart, glands in the digestive canal, sweat glands, and adrenal and other endocrine glands. Autonomic neuropathy is relatively uncommon but may occasionally occur in occupational intoxication, giving rise to bowel and bladder disturbance and postural hypertension.[3] Symptoms of autonomic nervous system dysfunction have been described in workers exposed to organic solvents and pesticides.[4,5] An outcome used to assess the existence of autonomic neuropathy is postural hypotension (also referred to as postural drop): a decrease in blood pressure upon standing from a seated or lying position. It can lead to feelings of faintness, light-headedness, weakness, unsteadiness, vertigo, poor concentration, headache or nausea. These were assessed in the current study using measured postural hypotension as well as several items from the Postal Questionnaire (i.e. feeling faint, dizzy, having blackouts, and feeling faint or fainting when standing up from a lying or sitting position).

2) The reported association between exposure to organic solvents and cardiac arrhythmias.[6] This was assessed by asking in the Postal Questionnaire about the occurrence of a rapid, pounding or irregular heartbeat.

3) The reported association between occupational and environmental toxins and the development, or acceleration, of heart disease.[1,7,8] This was assessed using several items from the Postal Questionnaire (i.e. experiencing chest pain, and being previously diagnosed by a medical practitioner as having heart disease or high blood pressure).

Postural hypotension was included as part of the General Health and Medical Study health examination as one measure of autonomic peripheral nerve dysfunction possibly arising from occupational exposure to F-111 DSRS chemicals. In addition, several self-report items from the Postal Questionnaire were chosen as indicative of cardiovascular health. These were: feeling faint, dizzy, having blackouts, feeling faint or fainting when standing up from a lying or sitting position, experiencing chest pain and/or a rapid, pounding or irregular heart beat, and

being previously diagnosed by a medical practitioner as having heart disease or high blood pressure. The use of medications for treatment of heart disease was also included in these analyses.

10.2 Measures

10.2.1 Health examination

Pulse and blood pressure were measured in three positions: seated position, lying, and standing. The lying and standing measures were used to assess postural hypotension. Blood pressure was measured twice in the seated position, with a five-minute interval between measures. After the second seated blood pressure measurement, the participant was asked to lie down. The test for vibration sensation (biothesiometry) was administered. Following this, with the participant still in the lying position, the lying pulse and blood pressure were taken. The participant was then asked to stand. A minimum two-minute (and no more than three-minute) delay was required before obtaining the standing pulse and blood pressure.

Blood pressure was taken on the right arm, using a Baumanometer mercury sphygmomanometer. The systolic pressure was taken as the first appearance of any sound (Korotkoff Phase I) and the diastolic pressure was taken as the complete disappearance of any sound (Korotkoff Phase V). Doctors were instructed to deflate the cuff at a rate of about 2mm per heartbeat. There also had to be a minimum of five minutes between the first and second seated blood pressure measurements. Pulse was taken at the right radial point of the wrist and counted over 60 seconds. Doctors were instructed to specify “regular” or “irregular” when recording results. Criteria for an indication of postural hypotension were a decrease in systolic blood pressure of 20 mm/Hg or a decrease in diastolic blood pressure of at least 10 mm/Hg,[9] between lying and standing measures, taken with a minimum of two minutes between each reading.

10.2.2 Postal Questionnaire items

Items taken from the Postal Questionnaire to assist the overall description of cardiovascular health were whether the participant reported experiencing in the past month “chest pain”

(item 2.1), “rapid, pounding or irregular heart beat” (item 2.3), “dizziness” (item 2.31), “fainting” (item 2.33), “blackouts” (item 2.34), or “feeling like you will faint, or fainting, when you stand up from a lying or sitting position” (item 2.79). Also, each person was asked to report if they had been diagnosed by a medical practitioner as having “high blood pressure” (item 2.81) and/or “heart disease” (item 2.82).

10.2.3 Medications

From the Postal Questionnaire, self-reported medication use was included as part of the analyses where any of the following types of medications were indicated:

Antihypertensives

- antiadrenergic agents, centrally acting
- antiadrenergic agents, ganglion-blocking
- antiadrenergic agents, peripherally acting
- arteriolar smooth muscle, agents acting on
- other antihypertensives
- antihypertensives and diuretics in combination
- combinations of antihypertensives

Beta blocking agents

- beta blocking agents
- beta blocking agents and thiazides
- beta blocking agents and other diuretics
- beta blocking agents, thiazides and other diuretics
- beta blocking agents and vasodilators
- beta blocking agents and other antihypertensives

Calcium channel blockers

Agents acting on the Renin-Angiotensin system

- ACE inhibitors
- ACE inhibitors, combinations
- angiotensin II antagonists, plain
- angiotensin II antagonists, combinations.

10.3 Potential confounders

Age, posting category, rank, civilian solvent exposure, BMI, beta blocker use, nitrate use, aspirin use, ACE inhibitor use, alcohol and smoking status were considered as potential confounders for cardiovascular health outcomes. HSA centre was considered as a potential confounder for outcomes obtained from the health examination.

10.4 Analyses

All analyses followed the outline described in Chapter 6 (Analysis). Participants were classified as having postural hypotension if the standing systolic blood pressure minus the lying systolic blood was 20 mm Hg or more, or if the standing diastolic blood pressure minus the lying diastolic blood was 10 mm Hg or more.

10.5 Results

Information on diagnosis of postural hypotension was obtained for 1531 participants who underwent a health and medical examination (seven individuals had missing data for this variable). All 1726 individuals who completed a Postal Questionnaire were included in analyses of postural hypotension-related symptoms, arrhythmias and cardiovascular disease risk factors and symptoms.

10.5.1 Basic description of cardiovascular health

The mean for seated systolic and diastolic blood pressures and pulse rate was slightly elevated for Amberley compared with the exposed and Richmond groups (see Table 10.1). The proportions of those with irregular pulse were similar across groups. The proportions with elevated blood pressure were not similar for the three groups in that there was a smaller proportion of the exposed group with elevated systolic and diastolic blood pressure (see Table 10.2).

Table 10.1 : Least squares adjusted means of blood pressure and pulse rates

Measure	Amberley	Richmond	Exposed	p-value
Seated systolic blood pressure	130.5	129.6	128.0	0.008
Seated diastolic blood pressure	83.2	82.0	81.4	0.013
Seated pulse rate	70.2	69.1	69.7	0.27

Table 10.2 : Proportion of blood pressure and pulse values out of normal range

Measure	Amberley N=403		Richmond N=515		Exposed N=613	
	N	***%***	**N**	***%***	**N**	***%***
Seated systolic blood pressure >140	66	*16*	76	*15*	58	*9.5*
Seated diastolic blood pressure >90	75	*19*	96	*19*	70	*11*
	N=404	***%***	**N=515**	***%***	**N=613**	***%***
Seated pulse rate >100	4	*1.0*	6	*1.2*	4	*0.65*
	N=397	***%***	**N=502**	***%***	**N=603**	***%***
Irregular pulse	5	*1.3*	5	*1.0*	6	*1.0*

10.5.2 Postural hypotension

There was no heterogeneity in the postural systolic or diastolic blood pressure change between HSA centres (p=0.44 and p=0.37 respectively). Table 10.3 shows the number and percentage of participants who had a diagnosis of postural hypotension.

Table 10.3 : Number and percentage of participants with postural hypotension outcomes by exposure group

Postural hypotension outcomes	Amberley		Richmond		Exposed	
	N=405	%	N=515	%	N=611	%
Diagnosis of postural hypotension (HE)*	7	*1.7*	7	*1.4*	4	*0.7*
	N=486	%	N=589	%	N=635	%
Self-reported fainting in the past month (PQ)**	5	*1.0*	4	*0.7*	8	*1.3*
	N=486	%	N=589	%	N=635	%
Self-reported blackouts in the past month (PQ)	5	*1.0*	5	*0.9*	12	*1.9*
	N=485	%	N=589	%	N=635	%
Self-reported dizziness in the past month (PQ)	73	*15*	93	*16*	207	*33*
	N=486	%	N=591	%	N=632	%
Self-reported feeling faint or fainting when standing in the past month (PQ)	62	*13*	74	*13*	155	*25*

** Health Examination*
*** Postal Questionnaire*

Only 18 participants had postural hypotension: seven (1.7%) in the Amberley comparison group, seven (1.4%) in the Richmond comparison group, and four (0.7%) in the exposed group. There were too few individuals with this outcome for any meaningful analyses. The results, when analysed as a continuous variable, indicated no difference between the three groups in either systolic (Table 10.4) or diastolic blood pressure (Table 10.5).

Table 10.4 : Postural hypotension, based on change in systolic blood pressure – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	-2.65	-3.20	-2.68
Standard Deviation	7.49	8.15	7.99
50th Percentile	-2.00	-2.00	-2.00
Lower Quartile	-8.00	-8.00	-8.00
Upper Quartile	2.00	2.00	2.00

Table 10.5 : Postural hypotension, based on change in diastolic blood pressure – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	-6.01	-6.23	-6.11
Standard Deviation	6.74	6.11	6.41
50th Percentile	-5.00	-6.00	-5.00
Lower Quartile	-10.00	-10.00	-10.00
Upper Quartile	-2.00	-2.00	-2.00

For all three groups there was a net rise from sitting to standing systolic blood pressure of between 2.6 and 3.2 mm mercury, and a net rise from sitting to standing diastolic blood pressure of between 6.0 and 6.2 mm mercury. The multivariate linear regression model showed no effect of group on systolic postural hypotension in the primary analysis (all exposed, p=0.97), or secondary analyses (Program 1, p=0.61; Program 2, p=0.95) (see Table 10.6).

Table 10.6 : Health examination postural hypotension systolic – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Estimate	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.05	-1.02	1.11	0.09	2	0.9718
	Richmond vs Exposed	0.12	-0.89	1.14			
Program 1	Amberley vs Exposed	0.57	-0.78	1.93	0.09	2	0.6125
	Richmond vs Exposed	0.63	-0.67	1.94			
Program 2	Amberley vs Exposed	0.04	-1.36	1.43	0.10	2	0.9509
	Richmond vs Exposed	0.18	-1.19	1.55			
Dose	Mild exposure vs Unexposed	0.64	-0.73	2.00	0.09	3	0.5150
	Moderate exposure vs Unexposed	-0.67	-1.99	0.66			
	Prolonged exposure vs Unexposed	0.00	-1.30	1.31			

Similarly there was no difference in diastolic postural hypotension for all exposed (p=0.97), Program 1 (p=0.94) or Program 2 (p=1.0) (see Table 10.7). Neither outcome showed a graded dose-response curve. All results were similar with the reduced models and with the robust standard error estimates.

Table 10.7 : Health examination postural hypotension diastolic – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Estimate	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.10	-0.76	0.97	0.08	2	0.9650
	Richmond vs Exposed	0.09	-0.73	0.91			
Program 1	Amberley vs Exposed	0.18	-0.88	1.25	0.08	2	0.9438
	Richmond vs Exposed	0.13	-0.90	1.15			
Program 2	Amberley vs Exposed	0.05	-1.11	1.20	0.09	2	0.9964
	Richmond vs Exposed	0.04	-1.09	1.18			
Dose	Mild exposure vs Unexposed	0.82	-0.28	1.93	0.08	3	0.0489
	Moderate exposure vs Unexposed	0.01	-1.06	1.08			
	Prolonged exposure vs Unexposed	-1.08	-2.14	-0.03			

Self-reported postural hypotension-related symptoms (experienced in the previous month) were shown in Table 10.3. Very few participants (approximately 1-2%) in all groups reported experiencing fainting or blackouts in the past month. Therefore no further analyses were conducted on these outcomes.

The proportion of participants reporting either dizziness or feeling faint when standing from a sitting or lying position, was higher for the exposed group than for both the Amberley and Richmond comparison groups. In all multiple regression analyses, the odds of self-reported dizziness in the past month were statistically significantly lower for both Amberley and Richmond comparison groups relative to the exposed group (see Table 10.8). The Amberley comparison group had 0.38 times the odds of self-reported dizziness relative to the exposed group (95% CI: 0.26, 0.54), while the Richmond comparison group had 0.43 times the odds relative to the exposed group (95% CI: 0.31, 0.59). The results were similar for secondary

analyses for both Program 1 (p<0.0001) and Program 2 (p<0.0001), and a dose-response relationship was demonstrated (p<0.0001).

Table 10.8 : Self-reported dizziness in the past month – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.38	0.26	0.54	2	41.07	<.0001
	Richmond vs Exposed	0.43	0.31	0.59			
Program 1	Amberley vs Exposed	0.34	0.22	0.53	2	28.05	<.0001
	Richmond vs Exposed	0.39	0.26	0.59			
Program 2	Amberley vs Exposed	0.35	0.23	0.55	2	25.20	<.0001
	Richmond vs Exposed	0.40	0.27	0.61			
Dose	Mild exposure vs Unexposed	2.08	1.36	3.18	3	44.06	<.0001
	Moderate exposure vs Unexposed	2.20	1.47	3.29			
	Prolonged exposure vs Unexposed	3.22	2.21	4.70			

The odds of feeling faint or fainting when standing from sitting or lying down, were statistically significantly lower for the Amberley comparison group relative to the exposed group (OR=0.48; 95% CI: 0.33, 0.70) and for the Richmond comparison group relative to the exposed group (OR=0.45; 95% CI: 0.32, 0.64) (see Table 10.9). This significant relationship remained in the secondary analyses for Program 1 and Program 2 (OR=0.40-0.44 for all these analyses). There was a moderate dose-response relationship with odds ratios of 1.85, 2.36 and 2.39 for low, medium and high exposure, relative to unexposed. These self-reported symptom results were also consistent in the reduced models.

Table 10.9 : Self-reported feeling faint or fainting when standing (from sitting or lying position) in the past month – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.48	0.33	0.70	2	26.06	<.0001
	Richmond vs Exposed	0.45	0.32	0.64			
Program 1	Amberley vs Exposed	0.41	0.26	0.64	2	21.66	<.0001
	Richmond vs Exposed	0.40	0.26	0.61			
Program 2	Amberley vs Exposed	0.44	0.28	0.70	2	17.41	0.0002
	Richmond vs Exposed	0.41	0.26	0.64			
Dose	Mild exposure vs Unexposed	1.85	1.18	2.91	3	27.54	<.0001
	Moderate exposure vs Unexposed	2.36	1.55	3.61			
	Prolonged exposure vs Unexposed	2.39	1.59	3.59			

10.5.3 Cardiovascular disease risk factors and symptoms

Less than 5% of participants reported a previous physician-diagnosis of heart disease, with a similar distribution across all groups (see Table 10.10). No further analyses were conducted on this outcome. Table 10.10 also shows the distribution of self-reported chest pain in the past month as well as previous physician diagnosis of high blood pressure.

Table 10.10 : Number and percentage of participants with cardiovascular disease outcomes by exposure group

Cardiovascular disease outcome	Amberley		Richmond		Exposed	
	N=478	***%***	**N=583**	***%***	**N=621**	***%***
Self-report of previous physician diagnosis of heart disease (PQ)	23	*4.8*	25	*4.3*	30	*4.8*
	N=480	***%***	**N=583**	***%***	**N=623**	***%***
Self-report of previous physician diagnosis of high blood pressure (PQ)	115	*24*	121	*21*	130	*21*
	N=486	***%***	**N=588**	***%***	**N=633**	***%***
Self-reported chest pain in the past month (PQ)	73	*15*	88	*15*	182	*29*
	N=485	***%***	**N=587**	***%***	**N=631**	***%***
Self-reported rapid, pounding or irregular heart beat in the past month (PQ)	72	*15*	90	*15*	156	*25*

The prevalence was higher in the exposed group than the Amberley and Richmond comparison groups for the two self-reported symptom outcomes. But it was similar (21-24%) for previous physician diagnosis of high blood pressure, where multivariate analysis indicated no group effect – neither overall, nor in subgroup analyses of Programs 1 and 2, nor in the dose-response curve (see Table 10.11).

Table 10.11 : Self-reported previous physician diagnosis of high blood pressure – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	1.23	0.82	1.84	2	1.04	0.5935
	Richmond vs Exposed	1.15	0.79	1.67			
Program 1	Amberley vs Exposed	1.15	0.70	1.88	2	0.39	0.8224
	Richmond vs Exposed	1.03	0.65	1.64			
Program 2	Amberley vs Exposed	1.55	0.87	2.79	2	2.32	0.3135
	Richmond vs Exposed	1.49	0.84	2.64			
Dose	Mild exposure vs Unexposed	0.80	0.46	1.38	3	1.18	0.7586
	Moderate exposure vs Unexposed	0.83	0.49	1.42			
	Prolonged exposure vs Unexposed	0.84	0.50	1.40			

In the multiple regression analyses, the odds of self-reported chest pain in the past month for both comparison groups was less than half that of the exposed group (OR=0.38, 95% CI 0.26, 0.55; OR=0.49, 95% CI: 0.35, 0.67 for Amberley and Richmond versus comparison group respectively) (see Table 10.12). The odds ratios were slightly lower for Program 1 (OR=0.31, 95% CI 0.20, 0.48; OR=0.41, 95% CI: 0.28, 0.61 for Amberley and Richmond versus comparison group respectively) and slightly higher for Program 2 (OR=0.44, 95% CI 0.27, 0.71; OR=0.57, 95% CI: 0.37, 0.89 for Amberley and Richmond respectively). A significant dose-response relationship was demonstrated, with OR of 1.78, 2.15 and 2.56 for low, medium and highest tertiles of exposure compared to no exposure. These results were similar in the reduced model.

Table 10.12 : Self-reported chest pain in the past month – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.38	0.26	0.55	2	33.12	<.0001
	Richmond vs Exposed	0.49	0.35	0.67			
Program 1	Amberley vs Exposed	0.31	0.20	0.48	2	29.70	<.0001
	Richmond vs Exposed	0.41	0.28	0.61			
Program 2	Amberley vs Exposed	0.44	0.27	0.71	2	11.71	0.0029
	Richmond vs Exposed	0.57	0.37	0.89			
Dose	Mild exposure vs Unexposed	1.78	1.15	2.77	3	28.83	<.0001
	Moderate exposure vs Unexposed	2.15	1.42	3.24			
	Prolonged exposure vs Unexposed	2.56	1.73	3.80			

10.5.4 Arrythmias

Table 10.13 shows that the exposed group reported more episodes of rapid, pounding or irregular heartbeat than the two comparison groups. Both comparison groups had statistically significantly lower odds of self-reported rapid, pounding or irregular heart beat in the past month than the exposed group for all analyses (OR of 0.45 for Amberley and 0.65 for Richmond, versus exposed) (see Table 10.13). For primary analysis and secondary, the odds ratios were between 0. 41 and 0.68. The dose-response analysis was also statistically significant, with odds ratios for low, medium and highest tertiles of exposed relative to unexposed of 1.36, 1.56 and 2.24. The reduced model gave similar results.

Table 10.13 : Self-reported rapid, pounding or irregular heartbeat in the past month – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.45	0.31	0.67	2	17.83	0.0001
	Richmond vs Exposed	0.65	0.47	0.90			
Program 1	Amberley vs Exposed	0.41	0.26	0.66	2	13.55	0.0011
	Richmond vs Exposed	0.61	0.40	0.93			
Program 2	Amberley vs Exposed	0.47	0.29	0.76	2	9.67	0.0079
	Richmond vs Exposed	0.68	0.44	1.06			
Dose	Mild exposure vs Unexposed	1.36	0.85	2.16	3	16.63	0.0008
	Moderate exposure vs Unexposed	1.56	1.01	2.41			
	Prolonged exposure vs Unexposed	2.24	1.50	3.34			

10.6 Discussion

Given that postural hypotension was a rare finding (approximately 1% of participants), the statistical power to detect a difference was low and no further analyses were done. Analysing this as a continuous measure indicated no difference between the three groups. However, self-reported dizziness or feeling faint when standing was significantly more common in the exposed group. The odds ratio for dizziness was 2.6 (95% CI 1.9-3.8) versus the Amberley group, and 2.3 (95% CI 1.7-3.2) versus Richmond. The odds ratio for feeling faint when standing was 2.1 (95% CI 1.4-3.0) versus Amberley, and 2.2 (95% CI 1.6-3.1) versus Richmond. Both of these symptoms were consistent and statistically significant in the Program 1 and 2 subgroups and both showed some suggestion of a dose-response, with that for dizziness being more convincing, i.e. odds ratios increasing from 2.0 to 2.2 to 3.2 from the lowest to highest exposed groups. Other cardiovascular end-points showed a similar pattern. There was no difference in physician-diagnosed heart disease or high blood pressure, but there was an increase in self-reported chest pain and palpitations. Chest pain was roughly twice as common in the exposed group (odds ratio 2.6 (95% CI 1.8-3.8) versus Amberley, and 2.0 (95% CI 1.5-2.9) versus Richmond), as were palpitations (OR=2.2 95% CI 1.5-3.2) versus Amberley and 1.5 (95% CI 1.1-2.1) versus Richmond. Both of these were consistent and significant in the Program 1 and 2 subgroup comparisons, and both showed a dose-response effect. The odds ratio for chest pain increased from 1.8 to 2.2 to 2.6 from the lowest to highest exposure groups, and for palpitations it increased from 1.4 to 1.6 to 2.2.

10.7 Conclusions

In conclusion there was a statistically significant increase in all self-reported cardiac symptoms from light-headedness to palpitations to chest pain. This was consistent in subgroup analyses and showed a dose-response effect. However, there were no differences found during the physical examination.

10.8 References

1. Benowitz, N.L. Chapter 20, Clinical Cardiac Toxicology. in *Clinical Environmental Health and Toxic Exposures* (eds. Sullivan, J.B. & Krieger, G.R.) 260-270 (Lippincott Williams & Wilkins, Philadelphia, 2001).
2. Kristensen, T.S. Cardiovascular diseases and the work environment. A critical review of the epidemiologic literature on chemical factors. *Scandinavian Journal of Work Environment & Health* **15**, 245-64 (1989).
3. Manzo, L. & Cosi, V. Manifestation of neurotoxicity in occupational diseases. in *Occupational Neurotoxicology* (ed. Manzo, L.) 1 (CRC Press, Boca Raton, 1998).
4. Juntunen, J., Hupli, V., Hernberg, S. & Luisto, M. Neurological picture of organic solvent poisoning in industry. A retrospective clinical study of 37 patients. *Int Arch Occup Environ Health* **46**, 219-31 (1980).
5. Murata, K., Araki, S., Yokoyama, K. & Maeda, K. Autonomic and peripheral nervous system dysfunction in workers exposed to mixed organic solvents. *Int Arch Occup Environ Health* **63**, 335-40 (1991).
6. Steinhauer, S.R., Morrow, L.A., Condray, R. & Scott, A.J. Respiratory sinus arrhythmia in persons with organic solvent exposure: comparisons with anxiety patients and controls. *Archives of Environmental Health* **56**, 175-80 (2001).
7. Hansen, E.S. International Commission for Protection Against Environmental Mutagens and Carcinogens. ICPEMC Working Paper 7/1/2. Shared risk factors for cancer and atherosclerosis--a review of the epidemiological evidence. [Review]. (1990).
8. Rosenman, K.D. Environmentally related disorders of the cardiovascular system. *Medical Clinics of North America* **74**, 361-75 (1990).
9. Bradley, J.G. & Davis, K.A. Orthostatic hypotension. *American Family Physician* **68**, 2393-2398 (2003).

This page has been left blank intentionally
for formatting purposes

11 Respiratory Health

Chapter summary

Respiratory health was assessed by spirometry testing, by previous physician diagnosis and by self-reported symptoms of chronic airway limitation. Based on pre- and post-Ventolin® lung function results, only five people were classified as having reactive airways disease (i.e. asthma-like symptoms). Self-reported physician diagnoses of bronchitis and emphysema (which together are clinically referred to as obstructive lung disease) were significantly elevated two-fold in the exposed group, and this was congruent with the two-fold elevation in self-reported symptoms of shortness of breath and wheezing in the exposed group versus the comparisons. This result is somewhat weakened by the lack of any difference in spirometry results ($FEV_1/FVC \leq 70\%$) at the health examination.

Chapter contents

11.1 Introduction

Respiratory symptoms were of interest in the current study for a number of reasons:

a) Nose and throat irritation, chronic cough and asthma-like conditions were highlighted in the Board of Inquiry (BOI).[1]

b) A number of DSRS workers registered respiratory complaints with the F-111 Interim Health Care Scheme.

c) Environmental exposures can cause both acute respiratory injury such as rhinosinusitis, laryngitis, upper airway obstruction, bronchitis, bronchoconstriction, alveolitis and pulmonary oedema, as well as chronic respiratory diseases such as asthma, bronchitis, parenchymal fibrosis and pleural fibrosis, and cancer.[2]

Numerous substances in the form of fibres, dust, vapour, aerosol, mist, fumes, smoke or gas may affect the function of the respiratory system. Of particular concern for SHOAMP were:

a) *Isocyanates*. Numerous studies report adverse respiratory effects of exposure to toluene diisocynate (TDI).[3-7] These compounds, used in the Spray Seal DSRS Program, are irritating to the mucous membranes and respiratory tract. Overexposure to TDI may cause sensitisation such that subsequent respiratory exposure may result in allergic rhinitis or allergic asthma.[8] Occupational asthma associated with TDI has five major components:
   - occupational asthma of the sensitiser type, which occurs in 5-10% of exposed workers weeks to months after the onset of exposure
   - chemical bronchitis
   - acute, but asymptomatic, deterioration of respiratory function during a work-shift
   - chronic deterioration of respiratory function associated with chronic exposure to low doses
   - persistent asthma or restrictive airway disease syndrome after exposure to high doses.[9]

b) *Organic solvents*. All organic solvents irritate the respiratory tract as a consequence of their de-fatting actions. In a recent review of the respiratory effects of organic solvents, Schenker[10] found that population-based epidemiological studies utilising job-exposure matrices have observed an independent association of solvent exposure and both respiratory symptoms and reduced pulmonary function.

c) *Jet Fuel and Jet Stream Exhaust*. A number of recent publications have investigated the respiratory system effects of jet fuel.[11-14] At RAAF Bases Amberley and Richmond, kerosene-based jet fuel vapour/aerosol and fuel combustion exhaust products enter the atmosphere from a wide range of sources. In addition, DSRS workers in the depuddling process experienced high exposures to jet fuel. Tunnicliff's[11] study found no increase in the prevalence of respiratory symptoms or change in spirometry in a sample of 222 UK airport workers with different levels of exposure to aircraft fuel and/or jet stream exhaust. In contrast, the Air Force Institute for Environment Safety and Occupational Health Risk Analysis (AFIERA) study of occupational JP-8 (fuel) exposure indicated a significant increase in self-reported difficulty with breathing and chest tightness among personnel exposed repeatedly to high-dose compared to low-dose concentrations.[15] The common thread through these studies, however, is the lack of long-term chronic effects.

11.2 Measures

11.2.1 Spirometry

The assessment of pulmonary function to detect and quantify abnormal lung function in epidemiological studies was based on spirometry: forced expiratory volume in one second (FEV_1), forced vital capacity (FVC), and the FEV_1/FVC ratio. These measures provide the best method of detecting the presence and severity of airway obstruction as well as the most reliable assessment of overall respiratory impairment (Table 11.1).

Table 11.1 : Measurements of pulmonary function

Measure	Significance
Forced expiratory volume at 1 second - FEV_1	The most common measure of airflow obstruction; estimates amount of air forced from lungs in 1 second of effort.
Forced vital capacity - FVC	The best estimate of the amount of air that can be exchanged in a single breath. Does not include residual volume left in lungs at end of expiration. Low FVC suggests either restriction or implies air trapping, with air left in lungs at end of expiration (see FEV_1/FVC ratio)
Forced expiratory volume at 1 second / forced vital capacity – FEV_1/FVC ratio	To understand the relevance of FEV_1, it must be considered in ratio to the total amount of expired air (see FVC). High to normal ratio suggests no airflow obstruction; low ratio indicates reduced airflow compared to total amount of air expirable.
Bronchodilator testing – Ventolin®	To assess the possibility of asthma FEV_1 & FVC, measures were repeated after the inhalation of a measured dose of the bronchodilator drug, Ventolin® (salbutamol sulfate, GlaxoWellcome). Improvement in spirometry test results indicating improvement in airways obstruction and the possibility of asthma.

The testing of lung function was performed according to American Thoracic Society criteria (see shaded box on the following page).[16] The SHOAMP measured respiratory function pre- and post-administration of Ventolin®, with three technically-satisfactory tests needed for each type; the participant could do up to, but no more than, eight blows until three were satisfactory. Each participant was required to sign an additional consent form prior to receiving Ventolin®. If they refused, only their pre-Ventolin® results were recorded. If they gave their consent, three post-bronchodilator tests were conducted. A metered-dose inhaler was used (Ventolin®, 100*ug* salbutamol/puff, Glaxo-Wellcome, Melbourne, Victoria Australia), and shaken for a few seconds before use. The participant was asked to exhale, then seal their lips around the inhaler; the inhaler was then activated and the participant was asked to take a slow maximal inhalation and then to hold their breath for approximately 5-10 seconds (providing it did not cause them discomfort). Following the administration of Ventolin®, lung function testing was repeated.

If a person was unable to perform the required number of tests in comfort, this was recorded in the Health Examination Booklet by the attending nurse (see Appendices 11A and 11B for set-up procedures for the Office Medic Spirocard and machine calibration).

Each participant was asked to breathe in fully (inspire) away from the mouthpiece, seal their lips tightly around the mouthpiece, and then blow the air out into the Spirometer as hard, fast and completely as they could until they were told to stop.

Spirometry results were entered into SQL server files from the Spirometry program, and were converted to MS Access for reading.

A number of medications were identified that could affect lung function testing. In the letter confirming their health examination appointment, participants received guidelines regarding the use of medications prior to their appointment. The ideal time intervals between last medication and lung function testing are shown in Table 11.2. If a participant had taken any of the listed medication types within the time period specified, this was indicated by the attending nurse who ticked the appropriate box in the Health Examination Booklet. The Lung Function Test was still performed.

ATS criteria for the forced expiratory manoeuvre technique

SHOAMP lung function testing was performed according to American Thoracic Society criteria:

- All forced expiratory manoeuvres had to be performed with the participant in a sitting position (refer to Figure 11.1).
- A new mouthpiece had to be used for each person and discarded at the completion of their session.
- The nurse made a nose clip available to participants (as is recommended), but its use was not compulsory.
- The forced expiratory manoeuvre had to be performed with maximum effort immediately following a maximum inspiration.
- The flow-volume curve had to be recorded without cough and have a clearly defined peak followed by a progressive decrease in expiratory flow down to zero flow.

A participant was asked to repeat the test if any of the following actions occurred:

- They failed to take a full inspiration.
- There was an unsatisfactory start to expiration (characterised by excessive hesitation or false start).
- They failed to expire for a minimum of 6 seconds.
- There was coughing during the first second of the manoeuvre which would affect the measurement of FEV_1 or there was any cough that interfered with the accurate measurement of FVC.
- There was evidence of valsalva manoeuvre (glottis closure), as indicated by truncation of the flow-volume curve.
- The participant had a leaky mouthpiece or there was obstruction of the mouthpiece.
- The participant failed to put the mouthpiece properly into their mouth.

Manoeuvres which had any of these faults had to be rejected by the attending nurse as a failed attempt, and the participant was encouraged to produce a better reading.

Figure 11.1 : Respiratory testing

Table 11.2 : Medications that could affect lung function testing

Drug	Description	Time interval	Examples
ß agonists	Inhaled, short acting	> 8 hours	Ventolin®, Asmol, Respax, Respolin (all Salbutamol), Bricanyl (Terbutaline), Berotec (Fenoterol), Alupent (Orciprenaline)
ß agonists	Inhaled, long acting	> 24 hours	Serevent (Salmeterol xinafoate).
ß agonists	Oral	> 12 hours	Bricanyl (Terbutaline), Ventolin® (Salbutamol)
Anticholinergics	Inhaled, short acting	> 8 hours	Atrovent (Ipratropium bromide)
Theophylline preps	Oral, short acting	> 8 hours	Brondecon
Theophylline preps	Oral, long acting	> 12 hours	Austyn, Nuelin, Theo-dur
Sodium cromoglycate / nedocromil sodium		> 24 hours	Tilade, Intal, Cromese sterinebs

11.2.2 Self-reported symptoms of chronic airway limitation

In addition to spirometry data collected by HSA nursing staff, each participant reported symptoms of chronic airway limitations. From the Postal Questionnaire, items 2.5 and 2.6 asked whether the person had experienced “shortness of breath” and/or “wheezing” respectively, item 2.88 asked whether they had ever been diagnosed with “asthma”, and items 2.89, 2.90, 2.91 and 2.92 each asked about previous physician diagnoses of “bronchitis”, “pneumonia”, “tuberculosis” and “emphysema” (which together are clinically referred to in the current study as obstructive lung disease). Each item required a yes or no response, and some items required the year of diagnosis by a medical practitioner and whether or not the person was still receiving treatment.

11.3 Potential confounders

Potential confounders of interest for respiratory outcomes include age, posting category, rank, civilian exposure to organic solvents, and smoking status. The latter is important because of the well-documented negative effect of smoking on respiratory function. For the airways disease outcomes, which were obtained from the spirometry measures as part of the General Health and Medical Study, HSA centre was also included as a potential confounder.

11.4 Analyses

All analyses followed the outline described in Chapter 6 (Analysis). Respiratory outcomes were considered as three groups:

- reactive
- obstructive airways disease diagnosis – based on spirometry data and self-reported previous physician diagnosis of respiratory condition (bronchitis, pneumonia, tuberculosis, emphysema, asthma and bronchitis or emphysema)
- self-reported symptoms of shortness of breath or wheezing in the past month.

A combined outcome of shortness of breath or wheezing within the past month was also obtained, and individuals were classified as having this outcome if they reported the

presence of either shortness of breath or wheezing. Participants were diagnosed as having reactive (or reversible) airways disease if the post-Ventolin® value of FEV_1/FVC minus the pre-Ventolin® value of FEV_1/FVC was greater than 15% (indicating an improvement of more than 15%). A diagnosis of obstructive airways disease was provided if the value of pre-Ventolin® FEV_1/FVC was less than or equal to 70%.

11.5 Results

Data on the presence or absence of obstructive airways disease were available for 1483 participants. Data was missing for 55 participants, who did not undertake testing. As numerous participants did not consent to the administration of Ventolin®, only 1295 participants were included in these analyses for reactive airways disease.

11.5.1 Diagnosed airways disease

As shown in Table 11.3, only five participants were classified as having reactive airways disease (from health examination testing), so no further analyses could be conducted.

Table 11.3 : Number and percentage of participants with reactive and obstructive airways disease by exposure group

Airways disease outcome	Amberley		Richmond		Exposed	
	N=349	%	N=430	%	N=516	%
Reactive airways disease (HE)	0	0	1	0.2	4	0.8
	N=385	%	N=503	%	N=595	%
Obstructive airways disease (HE)	27	7.0	28	5.6	45	7.6

The proportion of participants with obstructive airways disease was similar for the three exposure groups (7.0%, 5.6% and 7.6% for Amberley comparison, Richmond comparison and exposed groups respectively). There was no heterogeneity across HSA centres

(p=0.292 for Amberley compared to exposed and p=0.133 for Richmond compared to exposed).

No statistically significant association between exposure group and the presence of obstructive airways disease was demonstrated in the primary analysis of all exposed (p=0.47), Program 1 (p=0.94) or Program 2 (p=0.36) subgroups (see Table 11.4). There was also no dose-response effect (p=0.16). These results were similar for the reduced model and for the robust standard error estimates.

Table 11.4 : Obstructive airways disease – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.74	0.42	1.29	2	1.52	0.4672
	Richmond vs Exposed	0.76	0.44	1.30			
Program 1	Amberley vs Exposed	1.04	0.53	2.06	2	0.13	0.9367
	Richmond vs Exposed	1.12	0.58	2.18			
Program 2	Amberley vs Exposed	0.60	0.29	1.25	2	2.06	0.3569
	Richmond vs Exposed	0.63	0.31	1.31			
Dose	Mild exposure vs Unexposed	1.28	0.63	2.63	3	5.20	0.1574
	Moderate exposure vs Unexposed	1.85	0.99	3.43			
	Prolonged exposure vs Unexposed	0.78	0.36	1.67			

11.5.2 Self-reported physician-diagnosed respiratory conditions

Table 11.5 shows the prevalence of previous physician diagnosed respiratory conditions by exposure group. Very few participants reported a diagnosis of tuberculosis or emphysema. Diagnosis of pneumonia was similar across the three groups (8.6%-10%), as was asthma (10%-13%). Physician diagnosis of bronchitis was similar for the two comparison groups (11% and 12%) and slightly higher for the exposed group (19%).

Table 11.5 : Number and percentage of participants with self-reported previous physician diagnosis of respiratory conditions by exposure group

Self-reported previous physician diagnosis of respiratory symptom	Amberley		Richmond		Exposed	
	N=478	%	N=582	%	N=624	%
Pneumonia	41	*8.6*	54	*9.3*	65	*10*
	N=479	%	N=583	%	N=624	%
Tuberculosis	2	*0.4*	4	*0.7*	4	*0.6*
	N=479	%	N=582	%	N=625	%
Asthma	50	*10*	76	*13*	83	*13*
	N=479	%	N=582	%	N=623	%
Bronchitis	55	*11*	72	*12*	121	*19*
	N=477	%	N=583	%	N=624	%
Emphysema	3	*0.6*	3	*0.5*	9	*1.4*
	N=479	%	N=583	%	N=625	%
Bronchitis/Emphysema	55	*11*	74	*13*	124	*20*

There was no statistically significant association between exposure group and self-reported previous physician diagnosis of asthma in the overall analysis (p=0.52) or in the subgroup analyses (p=0.57 for Program 1, p=0.22 for Program 2) (see Table 11.6). These results were consistent with the reduced model. Given that in clinical practice bronchitis and emphysema are combined under the heading of chronic obstructive lung disease, these two were analysed in combination. There was a strong decrease in bronchitis/emphysema for both the Amberley comparisons (OR=0.48; 95% CI 0.33, 0.69) and the Richmond comparisons (OR=0.55; 95% CI 0.40, 0.77) versus all exposed (Table 11.7). This was consistent in both Program 1 (p=0.01) and Program 2 (p=0.001) subgroups and in the reduced model. Although

the dose effect was significant, there was no stepwise increase in risk with increasing involvement in F-111 DSRS activities, with the odds ratio ranging from 1.57 to 2.76 to 1.59.

Table 11.6 : Self-reported previous physician diagnosis of asthma – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.81	0.55	1.21	2	1.31	0.5201
	Richmond vs Exposed	1.00	0.70	1.43			
Program 1	Amberley vs Exposed	1.00	0.60	1.67	2	1.11	0.5747
	Richmond vs Exposed	1.20	0.75	1.92			
Program 2	Amberley vs Exposed	0.65	0.40	1.06	2	3.01	0.2221
	Richmond vs Exposed	0.77	0.49	1.22			
Dose	Mild exposure vs Unexposed	0.76	0.43	1.35	3	3.75	0.2894
	Moderate exposure vs Unexposed	1.42	0.90	2.24			
	Prolonged exposure vs Unexposed	1.05	0.65	1.68			

Table 11.7 : Self-reported previous physician diagnosis of bronchitis/emphysema – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.48	0.33	0.69	2	20.29	<.0001
	Richmond vs Exposed	0.55	0.40	0.77			
Program 1	Amberley vs Exposed	0.52	0.34	0.82	2	9.14	0.0104
	Richmond vs Exposed	0.60	0.40	0.90			
Program 2	Amberley vs Exposed	0.43	0.27	0.68	2	13.85	0.0010
	Richmond vs Exposed	0.51	0.33	0.78			
Dose	Mild exposure vs Unexposed	1.57	0.98	2.50	3	26.22	<.0001
	Moderate exposure vs Unexposed	2.76	1.85	4.11			
	Prolonged exposure vs Unexposed	1.59	1.03	2.44			

11.5.3 Self-reported respiratory symptoms in the past month

The prevalence of self-reported respiratory symptoms was higher in the exposed group than in both the Amberley and Richmond comparison groups for shortness of breath, wheezing, and the combined outcome of shortness of breath or wheezing (see Table 11.8).

Table 11.8 : Number and percentage of participants with self-reported respiratory symptoms in the past month by exposure group

Self-reported respiratory symptom	Amberley		Richmond		Exposed	
	N=487	***%***	**N=587**	***%***	**N=634**	***%***
Shortness of breath	135	*28*	136	*23*	255	*40*
	N=484	***%***	**N=587**	***%***	**N=635**	***%***
Wheezing	82	*17*	89	*15*	154	*24*
	N=487	***%***	**N=589**	***%***	**N=635**	***%***
Shortness of breath or wheezing	155	*32*	162	*28*	289	*46*

The multiple regression analyses demonstrated that the odds of shortness of breath or wheezing in the past month were statistically significantly lower for both comparison groups relative to the exposed group (see Table 11.9).

The odds of shortness of breath or wheezing in the past month, relative to the exposed group, were 0.58 for the Amberley comparison group (95% CI: 0.44, 0.76) and 0.50 for the Richmond comparison group (95% CI: 0.39, 0.65). These results were consistent with the Program 1 and Program 2 subgroup analyses (p=0.0001 and p=0.0042 respectively). A moderate dose-response relationship was demonstrated, with odds ratios of 1.52, 1.99 and 2.01 for mild, moderate and prolonged exposure compared to no exposure.

Table 11.9 : Self-reported shortness of breath or wheezing in the past month – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.58	0.44	0.76	2	31.05	<.0001
	Richmond vs Exposed	0.50	0.39	0.65			
Program 1	Amberley vs Exposed	0.52	0.37	0.73	2	24.10	<.0001
	Richmond vs Exposed	0.46	0.33	0.63			
Program 2	Amberley vs Exposed	0.63	0.44	0.91	2	10.95	0.0042
	Richmond vs Exposed	0.55	0.39	0.79			
Dose	Mild exposure vs Unexposed	1.52	1.06	2.18	3	28.80	<.0001
	Moderate exposure vs Unexposed	1.99	1.42	2.80			
	Prolonged exposure vs Unexposed	2.01	1.45	2.79			

11.6 Discussion

There was no increase in self-reported physician diagnosed asthma, and there were too few participants to analyse who had asthma-like results on spirometry (i.e. airway reactivity pre- and post-Ventolin®). In contrast, there was a clear increase in self-reported respiratory symptoms (shortness of breath or wheezing).

There was roughly a two-fold increase in the odds of these respiratory symptoms in the exposed group: a 1.7 fold increase (95% CI 1.3, 2.3) versus the Amberley comparison group and a 2.0 fold increase (95% CI 1.5, 2.6) versus Richmond. These results are consistent and remain significant for both Program 1 and Program 2 subgroups with some evidence of a dose-response effect.

These findings are supported by the results of previous physician diagnosed bronchitis/emphysema. There was a 2.1 fold increase in obstructive lung disease (95% CI 1.4, 3.0) compared to the Amberley group, and a 1.8 fold increase (95% CI 1.3, 2.5) compared to the Richmond group; this was consistent in Programs 1 and 2. The strength of these results, however, is somewhat diminished by the lack of any difference in results at the health examination; spirometry (FEV_1/FVC ratio $\leq$ 70%) did not detect any difference between the three groups, although the incidence of abnormal results was low. For example, although 20% of study participants from the exposed group reported a previous physician diagnosis of obstructive lung disease, only 7.6% had evidence of this on the day of examination.

11.7 Conclusions

In summary, there was no apparent association between exposure and asthma, although there was an increase in self-reported respiratory symptoms and physician diagnosed obstructive lung disease in the exposed group relative to the Amberley and Richmond comparison groups. The impact of this is somewhat lessened by the lack of any significant differences in spirometry measured at the physical examination.

11.8 References

1. Donaldson, E. Nature and extent of health complaints, Report to the F111 Deseal/Reseal Board of Inquiry. Volume 2, Part 1, Chapter 13, Annex B. (RAAF, Amberley, 2000).
2. Balmes, J.R. Occupational lung disease. in *Occupational and Environmental Medicine* (ed. LaDou, J.) 305-327 (Lange Medical Books/McGraw-Hill, New York, 1997).
3. Paggiaro, P.L. et al. Mechanisms of bronchial hyperresponsiveness in occupational asthma due to isocyanates. *Progress in Respiratory Research* **24**, 127-136 (1990).
4. Meredith, S.K., Bugler, J. & Clark, R.L. Isocyanate exposure and occupational asthma: a case-referent study. *Occuational and Environmental Medicine* **57**, 830-836 (2000).
5. Paggiaro, P.L. et al. TDI-induced oculorhinitis and bronchial asthma. *Journal of Occupational Medicine* **27**, 51-2 (1985).
6. Karol, M.H. & Hauth, B.A. Use of hexyl isocyanate antigen to detect antibodies to hexamethylene diisocyanate (HDI) in sensitized guinea pigs and in a sensitized worker. *Fundamentals of Applied Toxicology* **2**, 108-13 (1982).
7. Zedda, S., Cirla, A., Aresini, G. & Sala, C. Occupational type test for the etiological diagnosis of asthma due to toluene di-isocyanate. *Respiration* **33**, 14-21 (1976).
8. Collins, M. Toxicology of toluene diisocyanate. *Applied Occupational and Environmental Hygiene* **17**, 846-855 (2002).
9. Blames, J.R. & Scannell, C.H. Occupational Lung Disease. in *Occupational & Environmental Medicine* (ed. LaDou, J.) 305-327 (Lange Medical Books/McGraw-Hill, New Youk, 1997).
10. Schenker, M.B. & Jacobs, J.A. Respiratory effects of organic solvent exposure. *Tubercle and Lung Disease* **77**, 4-18 (1996).
11. Tunnicliffe, W.S. et al. Pulmonary function and respiratory symptoms in a population of airport workers. *Occupational and Environmental Medicine* **56**, 118-123 (1999).
12. Ritchie, G. et al. Biological and health effects of exposure to kerosene-based jet fuels and performance additives. *Journal Toxicology and Environmental Health, Part B Critical Reviews* **6**, 357-451 (2003).
13. Subcommittee on jet-propulsion fuel 8. Carcinogenic effects of jet-propulsion fuel 8. in *Toxicologic Assessment of Jet-Propulsion Fuel 8* (ed. National Research Council of the National Academies) 141-161 (The National Academies Press, Washington, D.C., 2003).

14. Wang, S., Young, R.S. & Witten, M.L. Age-related differences in pulmonary inflammatory responses to JP-8 jet fuel aerosol inhalation. *Toxicol Ind Health* **17**, 23-9 (2001).
15. Air Force Institute for Environment Safety and Occupational Health Risk Analysis. JP-8 final risk assessment. (Texas Tech University, Luddock, TX, 2001).
16. American Thoracic Society. Lung function testing: Selection of reference values and interpretative strategies. *American Review of Respiratory Diseases* **144**, 1202-1218 (1991).

12 Dermatological and Breast Abnormalities

Chapter summary

Complaints of adverse skin conditions by Deseal/Reseal (DSRS) workers prompted the inclusion in the General Health and Medical Study of a full skin examination to check for lipoma, dermatitis, psoriasis and “other” skin lesions, and of self-reported data about skin irritation, dermatitis, eczema, psoriasis and previously-diagnosed malignant melanoma. The issue of breast abnormalities was also of concern, as a male DSRS worker had previously been diagnosed with breast carcinoma and a second male had reported breast enlargement. There was a strong and statistically significant two- to three-fold increase in dermatitis in the F-111 DSRS group, and this was consistent between the different methods of assessment. This effect was more marked in comparison to the Amberley group than the Richmond group. There was a less robust two-fold increase in pigmented or sun-related lesions in the DSRS group compared to both the Amberley and Richmond groups. Other outcomes were either too rare or too variable to be analysed or they showed no difference between groups.

Chapter contents

12.1 Introduction

Dermatological outcomes were included in the current study for a variety of reasons:

a) Organic solvents are among the leading causes of occupational skin disease because they act as degreasers of human skin. This leads to the loss of the normal oils that protect skin from acute and chronic water loss, and to the development of chronic eczema, thickening, cracking, and drying.[1] With its protective layer dissolved, the skin is easily penetrated by other compounds.[2] The solvent methyl ethyl ketone (MEK) was widely used in all the Deseal/Reseal (DSRS) programs; of particular concern, however, is the reported use of MEK for cleaning "goop" from the hands, because this could enhance the penetration of other compounds which may cause systemic toxicity. Similarly, it has been recognised that military aircraft maintenance personnel are often unprotected against dermal exposure to kerosene-based jet fuels.[3] Ritchie reported that this reduction in the integrity of the dermal barrier by repeated exposure to kerosene-based jet fuel may cause itching or burning skin, skin redness or rash, skin dryness or dermatitis, skin lesions or weeping, or skin sensitisation; and it may increase systemic exposure to other occupational toxicants and environmental microbials, as well as to toxic components of the fuel itself during subsequent exposures.

b) Epoxy resins used in the DSRS process are well-known skin sensitisers. A workers' compensation claim study of occupational skin diseases in Washington State found that a high rate of claims was for contact dermatitis where employees are likely to have significant dermal exposure to epoxy and related resin systems.[4] Numerous other studies have reported occupationally-related contact dermatitis,[5-8] including positive findings from the aircraft manufacturing industry.

c) There were a number of F-111 DSRS workers who had registered complaints of adverse skin conditions with the Department of Veterans' Affairs (DVA) F-111 Interim Health Care Scheme.

The actual outcomes selected included:

a) *Lipomata*. Although the prevalence of lipomata was assessed in SHOAMP, the actual cause of these usually small fatty lumps under the skin remains uncertain. Evidence in

the scientific literature relating lipomata to occupational exposure is most often case reports or case studies involving only one person or a small number of subjects. One such study suggests a possible link to solvent exposure.[9]

b) *Skin cancers or potentially pre-malignant lesions*. In Australia, basal cell carcinoma and squamous cell carcinoma are generally thought to have their origins in excessive sun exposure; however, a number of occupational exposures have also been associated with these cancers. Arsenic, ionising radiation, mineral and shale oils, soots and tars all have well-documented associations with skin cancer, as have the industrial processes of coal gasification and the production of rubber.[10] In his recent review of the biological and health effects of exposure to kerosene-based jet fuels, Ritchie[3] presented a number of studies documenting the tumorigenic potential of repeated dermal exposure to fuels, which appear consistent with human data showing high incidence of dermatoses in ball-bearing workers exposed repeatedly to kerosene.

c) *Breast cancer*. The issue of breast cancer was relevant for inclusion in the SHOAMP General Health and Medical Study because one F-111 DSRS male worker had already been diagnosed with breast carcinoma and a second male had reported breast enlargement. It is hypothesised that organic solvents act directly as genotoxic agents, migrating to adipose tissue of the breast and remaining in a stored state, potentially initiating and promoting carcinogenesis.[11] A study by Blair et al.[12] evaluated cancer risk from potential exposure to trichloroethylene and other chemicals in a cohort of 14,457 aircraft maintenance workers over the period 1952-1990. Workers exposed to trichloroethylene showed excesses of breast cancer (RR 1.8), although this result was not statistically significant. An increased risk of breast cancer in men has also been associated with military service.[13] A number of longitudinal studies have also reported positive trends for breast cancer following exposure to organic solvents.[14,15]

d) Other lesions reported to the F-111 DSRS Interim Health Care Scheme included rash, itchiness, dryness and peeling skin.

12.2 Measures

12.2.1 Skin examination

Data on adverse skin conditions were collected via a visual skin examination conducted by HSA clinicians, as well as via separate self-reported items in the Postal Questionnaire. Visual examination by HSA clinicians was determined as the most effective method of identifying skin problems, given the amount of time available for all health tests to be conducted and the nature of the skin problems that were of specific interest to the Health Study. The skin examination identified the presence or absence of psoriasis, dermatitis, lipomata (defined as sub-cutaneous nodules, firm and smooth on palpation) and “other” skin neoplasms in participants. Where psoriasis or dermatitis was indicated, the clinician recorded the sites where they were present: for psoriasis – scalp, face, back, elbows, knees, nails or other; for dermatitis – hands, elbows, forearms, head/neck, knees, trunk or other. Handwritten comments regarding the presence and location of squamous cell and basal cell carcinoma, malignant melanoma, and “other skin lesion”, were also included as part of the analysis.

For the purposes of the skin examination, each participant was asked to remove their outer clothing (leaving on underwear). The participant was then instructed to lie on their back on the examination bed/couch with their palms facing up. Physicians were instructed to carefully examine the skin by sections, recording their observations in the Health Examination Booklet. Participants were then instructed to lie on their front with palms facing down, and the examination was repeated. In addition to the physical examination, the following self-reported symptoms from the Postal Questionnaire were also analysed: “rash or skin irritation” (item 2.13), physician-diagnosed “dermatitis” (item 2.114), “eczema” (item 2.115), “psoriasis” (item 2.115), and “malignant melanoma” (item 1.116).

12.2.2 Breast examination

Self-reported information on breast abnormalities was obtained from the Postal Questionnaire. Items 2.25-2.28 asked respondents to indicate whether or not they had experienced “any new lump(s) in the breast area”, “any change to the skin of the nipple/breast”, “an unusual increase in the size of one breast” and/or “sticky or bloody discharge from one/both nipples”. A breast examination was included as part of the SHOAMP health examination for both male and female participants. Each breast was

examined systematically, starting with the outer upper quadrant, first with the participant in a seated position, then lying down. In the Health Examination Booklet the attending doctor recorded any abnormalities as being either present or absent, providing more detail as required. The presence of gynaecomastia – enlargement of breast tissue in men, as defined by breast tissue extending beyond the areola – was also recorded.

12.3 Potential confounders

Potential confounders for all skin outcomes included age, posting category, and rank. Smoking status was included in analyses of melanoma and skin cancer. The general potential confounders considered for analyses of breast abnormalities were age group, rank, posting period, smoking, and alcohol. In addition, BMI was considered as a potential confounder because diagnosis of breast abnormalities can be difficult in the presence of fatty tissue. HSA centre was considered as a potential confounder for all outcomes assessed during the health examination.

12.4 Analyses

All analyses followed the outline described in Chapter 6 (Analysis). Several outcomes of interest from the Postal Questionnaire and the HSA health examination were considered for the skin analyses:

From the HSA health examination

- the presence of any physician-diagnosed lipoma
- the number of lipomata
- the total surface area of all lipomata
- physician diagnosis of psoriasis; dermatitis; melanoma, squamous cell carcinoma or basal cell carcinoma; and any pigmented or sun-related lesion.

From the Postal Questionnaire

- self-reported previous physician diagnosis of psoriasis, dermatitis, eczema, and malignant melanoma
- self-reported skin rash, irritation and skin ulcer in the past month.

12.5 Results

12.5.1 Skin examination

Lipomata

Lipomata were assessed during the health examination for 1532 participants. Of these, 121 people were identified as having one or more lipomata (see Table 12.1). The frequency of lipomata was similar among the three groups, with 8.1% of the exposed group having at least one lipoma as compared with 8.6% and 7.0% for the Amberley and Richmond comparison groups respectively (Table 12.1). The number of lipomata ranged from one to 40, with the next lowest count being 15 for two participants. The distribution of the number of lipomata was similar for the three exposure groups. The surface area of lipomata was estimated for each participant by summing the surface area of each lipoma counted. The person with 40 lipomata did not have lipoma length and breadth recorded. Surface area ranged between 0.01 cm^2 and 434 cm^2, with the next smallest area being 99 cm^2. To ensure a reasonably normal distribution of the residual values calculated in the regression, lipoma surface area was logged.

However, the Breslow Day test for homogeneity of odds ratios indicated that the relationship between exposure group and lipomata varied among centres (p=0.035 for Amberley compared to exposed, and p=0.008 for Richmond compared to exposed). Therefore it is not appropriate to obtain an overall measure of the association between exposure group and lipomata, and no further analyses were undertaken.

Table 12.1 : Number and percentage of participants with physician diagnosed lipomata, number of lipomata and surface area of lipomata by exposure group

Measure of Lipomata	Amberley N=406		Richmond N=516		Exposed N=615	
	n	***%***	**n**	***%***	**n**	***%***
Any physician diagnosed lipoma	35	*8.6*	36	*7.0*	50	*8.1*
Number of lipomata						
1	24	*69*	17	*50*	26	*53*
2	6	*17*	4	*12*	7	*14*
3	1	*3.0*	3	*9*	5	*10*
4+	4	*11*	10	*29*	11	*23*
Total	35	*100%*	34	*100%*	49	*100%*
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Surface area of all lipomata (cm^2)	13.28	21.7	18.82	27.8	22.5	65.1

Psoriasis

There was no heterogeneity between centres for physician diagnosis of psoriasis at the health examination (p=0.45 for Amberley compared to exposed, and p=0.16 for Richmond compared to exposed). Physician-diagnosed psoriasis at the time of the health examination was rare (<5%) in both the exposed and Richmond groups, and this was not analysed further due to a lack of statistical power (Table 12.2). The distribution of self-reported previous physician diagnosis of psoriasis appears similar across the three exposure groups. There was no statistically significant association between exposure group and self-reported previous physician diagnosis of psoriasis for the overall exposed group (p=0.31) or for Program 2 (p=0.73), although it was significant for Program 2 for Richmond versus exposed (OR=0.50, 95% CI 0.29, 0.88) (see Table 12.3). Although the dose-effect was significant (p=0.01), there was no clear stepwise increase in risk with increasing dose. The reduced model showed no group effect.

Table 12.2 : Number and percentage of participants with self-reported and physician diagnosed psoriasis by exposure group

	Amberley		Richmond		Exposed	
	N=475	**%**	**N=581**	**%**	**N=626**	**%**
Self-report of previous physician diagnosed psoriasis (PQ)	29	*6.1*	28	*4.8*	40	*6.4*
	N=405	**%**	**N=516**	**%**	**N=613**	**%**
Physician diagnosed psoriasis (HE)	22	*5.4*	25	*4.8*	18	*2.9*

Table 12.3 : Self-reported physician diagnosis of psoriasis – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.84	0.50	1.44	2	2.35	0.3092
	Richmond vs Exposed	0.67	0.40	1.12	.	.	.
Program 1	Amberley vs Exposed	0.62	0.34	1.11	2	5.91	0.0521
	Richmond vs Exposed	0.50	0.29	0.88	.	.	.
Program 2	Amberley vs Exposed	1.07	0.49	2.31	2	0.64	0.7252
	Richmond vs Exposed	0.85	0.40	1.83	.	.	.
Dose	Mild exposure vs Unexposed	1.15	0.55	2.41	3	10.69	0.0135
	Moderate exposure vs Unexposed	0.73	0.33	1.66	.	.	.
	Prolonged exposure vs Unexposed	2.36	1.35	4.15	.	.	.

Dermatitis

Table 12.4 shows the distribution of self-reported previous physician diagnosis of dermatitis and health examination physician diagnosis of dermatitis. For both of these outcomes, the prevalence appears higher in the exposed group than in either of the comparison groups, and the difference appears greater for self-reported previous physician diagnosis.

Table 12.4 : Number and percentage of participants with self-reported and physician diagnosed dermatitis by exposure group

	Amberley		Richmond		Exposed	
	N=477	***%***	**N=581**	***%***	**N=523**	***%***
Self-report of previous physician diagnosed dermatitis (PQ)	58	*12*	109	*19*	193	*31*
	N=403	***%***	**N=515**	***%***	**N=610**	***%***
Physician diagnosed dermatitis (HE)	46	*11*	67	*13*	97	*16*

As shown in Table 12.5, both Amberley and Richmond comparison groups had statistically significantly lower odds of self-reported previous physician diagnosis of dermatitis than the exposed group, with OR of 0.28 for Amberley compared to all exposed (95% CI 0.20, 0.40), and OR of 0.48 for Richmond compared to all exposed (95%CI 0.36, 0.63). This relationship was consistent in the Program 1 and 2 subgroups and in the reduced model. Although dose was statistically significant ($p<0.0001$), there was no clear stepwise increase in the risk with increasing dose.

Table 12.5 : Self-reported physician diagnosed dermatitis – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.28	0.20	0.40	2	58.88	<.0001
	Richmond vs Exposed	0.48	0.36	0.63	.	.	.
Program 1	Amberley vs Exposed	0.24	0.16	0.36	2	50.15	<.0001
	Richmond vs Exposed	0.41	0.29	0.57	.	.	.
Program 2	Amberley vs Exposed	0.24	0.16	0.37	2	43.55	<.0001
	Richmond vs Exposed	0.41	0.28	0.59	.	.	.
Dose	Mild exposure vs Unexposed	2.20	1.48	3.27	3	57.92	<.0001
	Moderate exposure vs Unexposed	3.14	2.19	4.52	.	.	.
	Prolonged exposure vs Unexposed	2.68	1.87	3.83	.	.	.

There was no heterogeneity between centres for Amberley versus exposed (p=0.86) or for Richmond versus exposed (p=0.87) for physician-diagnosed dermatitis at the time of the health examination. While group effect was not formally significant for physician diagnosis of dermatitis at the health examination for all exposed (p=0.06), Program 1 (p=0.32) or Program 2 (p=0.09), the point estimates indicated less dermatitis in the comparison groups (see Table 12.6). This was significant for the Amberley versus exposed comparison overall (OR 0.64; 95% CI 0.42, 0.96) and in Program 2 (OR 0.56; 95% CI 0.33, 0.96). This level of significance was increased in the reduced model and increased even further with the robust standard error estimates in which the Richmond versus exposed comparison also became significant (0.70; 95% CI 0.57, 0.85).

Table 12.6 : Health examination physician diagnosed dermatitis – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.64	0.42	0.96	2	5.60	0.0607
	Richmond vs Exposed	0.70	0.48	1.02	.	.	.
Program 1	Amberley vs Exposed	0.69	0.42	1.13	2	2.29	0.3184
	Richmond vs Exposed	0.77	0.49	1.22	.	.	.
Program 2	Amberley vs Exposed	0.56	0.33	0.96	2	4.74	0.0934
	Richmond vs Exposed	0.62	0.37	1.04	.	.	.
Dose	Mild exposure vs Unexposed	1.05	0.61	1.80	3	5.74	0.1248
	Moderate exposure vs Unexposed	1.52	0.95	2.45	.	.	.
	Prolonged exposure vs Unexposed	1.61	1.00	2.59	.	.	.

Other skin conditions

Table 12.7 shows the number and proportion of participants, by exposure group, with self-reported skin rash or irritation in the past month, self-reported skin ulcer within the past month, and self-reported previous physician diagnosis of eczema. The prevalence of self-reported skin rash or irritation in the past month seemed much higher in the exposed group than in both comparison groups. The distribution of self-reported skin ulcer in the past month and self-reported previous physician diagnosis of eczema appeared similar across the three exposure groups. As shown in Table 12.8, both Amberley and Richmond comparison groups had statistically significantly lower odds of self-reported skin rash or irritation in primary and secondary analyses.

Table 12.7 : Number and percentage of participants with other skin conditions by exposure group

	Amberley		Richmond		Exposed	
	N=486	***%***	**N=589**	***%***	**N=634**	***%***
Self-report of skin rash / irritation within the past month (PQ)	145	*30*	196	*33*	328	*52*
	N=486	***%***	**N=588**	***%***	**N=633**	***%***
Self-report of skin ulcer within the past month (PQ)	23	*4.7*	21	*3.6*	45	*7.1*
	N=477	***%***	**N=582**	***%***	**N=625**	***%***
Self-report of previous physician diagnosis of eczema (PQ)	22	*4.6*	27	*4.6*	36	*5.8*

Table 12.8 : Self-reported skin rash or irritation – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.39	0.30	0.51	2	63.51	<.0001
	Richmond vs Exposed	0.43	0.34	0.55	.	.	.
Program 1	Amberley vs Exposed	0.31	0.22	0.43	2	57.23	<.0001
	Richmond vs Exposed	0.34	0.25	0.47	.	.	.
Program 2	Amberley vs Exposed	0.43	0.30	0.61	2	23.75	<.0001
	Richmond vs Exposed	0.49	0.35	0.68	.	.	.
Dose	Mild exposure vs Unexposed	2.12	1.50	3.00	3	63.05	<.0001
	Moderate exposure vs Unexposed	2.43	1.74	3.38	.	.	.
	Prolonged exposure vs Unexposed	2.82	2.05	3.89	.	.	.

The odds were approximately 0.4 for both comparison groups for the primary analyses (OR for Amberley relative to exposed 0.39, 95% CI: 0.30, 0.51; OR for Richmond relative to exposed 0.43, 95% CI: 0.34, 0.55) and were consistent in the Program 1 and 2 subgroup analyses. A clear dose-response effect was present with odds ratios increasing from 2.1 to 2.4 to 2.8 with increasing exposure. Group effect was significant for self-reported skin ulcers in the past month for all exposed (p=0.03) (see Table 12.9).

Table 12.9 : Self-reported skin ulcer – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.67	0.38	1.17	2	6.93	0.0312
	Richmond vs Exposed	0.47	0.27	0.83	.	.	.
Program 1	Amberley vs Exposed	0.71	0.37	1.38	2	4.10	0.1290
	Richmond vs Exposed	0.51	0.26	0.98	.	.	.
Program 2	Amberley vs Exposed	0.50	0.24	1.02	2	7.95	0.0188
	Richmond vs Exposed	0.35	0.17	0.73	.	.	.
Dose	Mild exposure vs Unexposed	0.86	0.35	2.11	3	10.13	0.0175
	Moderate exposure vs Unexposed	2.31	1.26	4.25	.	.	.
	Prolonged exposure vs Unexposed	1.92	1.00	3.65	.	.	.

This was driven mostly by the Richmond versus exposed comparison (OR=0.47; 95% CI 0.27, 0.83) and remained significant in both the Program 1 and 2 subgroups and in the reduced model. Dose was statistically significant overall, but there was not a monotonically-increasing relationship between dose and the odds of skin ulcer. In multivariate analyses there was no significant relationship between self-reported previous physician diagnosis of eczema and exposure group (Table 12.10).

Table 12.10 : Self-reported previous physician diagnosis of eczema – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.66	0.37	1.19	2	2.40	0.3015
	Richmond vs Exposed	0.71	0.41	1.23	.	.	.
Program 1	Amberley vs Exposed	0.61	0.29	1.27	2	1.96	0.3752
	Richmond vs Exposed	0.66	0.34	1.31	.	.	.
Program 2	Amberley vs Exposed	0.72	0.33	1.59	2	0.67	0.7171
	Richmond vs Exposed	0.79	0.37	1.70	.	.	.
Dose	Mild exposure vs Unexposed	1.36	0.63	2.91	3	1.06	0.7859
	Moderate exposure vs Unexposed	1.25	0.59	2.67	.	.	.
	Prolonged exposure vs Unexposed	1.31	0.64	2.68	.	.	.

Cancerous lesions

Cancerous lesions included 24 cases of squamous cell carcinoma, 57 cases of basal cell carcinoma, and 24 cases of malignant melanoma. These were combined to form a group of 95 cases of cancerous lesions. There was no heterogeneity across HSA examination centres for this outcome (p=0.68 for Amberley compared to exposed, and p=0.13 for Richmond compared to exposed). On univariate analyses there appeared to be no association between self-reported melanoma diagnosis or physician diagnosis of cancerous lesions and exposure group (see Table 12.11). Multiple logistic regression analyses demonstrated no statistically significant association between self-reported physician diagnosis of skin cancer and group in the overall analysis (p=0.10, see Table 12.12), although the Amberley versus exposed comparison was significant (OR=0.50; 95% CI 0.26, 0.96) and remained so in the Program 2

subgroup (OR=0.34; 95% CI 0.15, 0.76) and the reduced model but not in Program 1 (OR=0.62; 95% CI 0.29, 1.34).

Table 12.11 : Number and percentage of participants with cancerous skin lesions by exposure group

	Amberley		Richmond		Exposed	
	N=477	***%***	**N=583**	***%***	**N=627**	***%***
Self-report of previous physician diagnosed melanoma (PQ)	14	*3.0*	33	*5.7*	40	*6.4*
	N=405	***%***	**N=516**	***%***	**N=613**	***%***
Physician diagnosed melanoma / SCC / BCC (HE)	27	*6.7*	33	*6.4*	41	*6.7*

Table 12.12 : Self-reported previous physician diagnosis of melanoma – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.50	0.26	0.96	2	4.59	0.1006
	Richmond vs Exposed	0.92	0.55	1.53	.	.	.
Program 1	Amberley vs Exposed	0.62	0.29	1.34	2	3.12	0.2104
	Richmond vs Exposed	1.12	0.59	2.11	.	.	.
Program 2	Amberley vs Exposed	0.34	0.15	0.76	2	7.08	0.0290
	Richmond vs Exposed	0.64	0.32	1.27	.	.	.
Dose	Mild exposure vs Unexposed	1.77	0.90	3.47	3	5.42	0.1437
	Moderate exposure vs Unexposed	1.49	0.76	2.94	.	.	.
	Prolonged exposure vs Unexposed	0.67	0.28	1.57	.	.	.

There was no dose-response effect seen. There was no association between group and skin cancer at the time of the health examination either in the overall comparison (p=0.62) or in the subgroups (Program 1 p=0.86, Program 2 p=0.71) (see Table 12.13).

Table 12.13 : Health examination physician diagnosed melanoma, SCC or BCC – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.85	0.49	1.49	2	0.94	0.6236
	Richmond vs Exposed	0.77	0.44	1.32	.	.	.
Program 1	Amberley vs Exposed	0.94	0.48	1.85	2	0.31	0.8568
	Richmond vs Exposed	0.84	0.43	1.62	.	.	.
Program 2	Amberley vs Exposed	0.84	0.39	1.81	2	0.68	0.7129
	Richmond vs Exposed	0.73	0.34	1.58	.	.	.
Dose	Mild exposure vs Unexposed	0.89	0.40	1.99	3	2.84	0.4169
	Moderate exposure vs Unexposed	1.66	0.87	3.18	.	.	.
	Prolonged exposure vs Unexposed	1.23	0.62	2.45	.	.	.

Pigmented or sun-related lesions

Given the ambiguity and uncertainty of making a diagnosis of skin cancer on clinical examination alone, the analysis was repeated using all pigmented or sun-related lesions (e.g. dysplastic or variegated moles, actinic or solar keratoses, etc). There was no heterogeneity across HSA examination centres (p=0.73 for Amberley compared to exposed, and p=0.77 for Richmond compared to exposed). Table 12.14 shows the distribution of these lesions by exposure group. The exposed group appeared to have a higher proportion of

lesions than both of the comparison groups. Table 12.15 refers to the multiple logistic regression analyses.

Table 12.14 : Number and percentage of participants with pigmented or sun-related lesions by exposure group

	Amberley		Richmond		Exposed	
	N=404	%	N=516	%	N=612	%
Physician diagnosed potentially cancerous skin lesions (HE)	87	*22*	120	*23*	190	*31*

Table 12.15 : Health examination physician diagnosed of pigmented or sun-related lesions – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.56	0.40	0.78	2	14.05	0.0009
	Richmond vs Exposed	0.64	0.47	0.87	.	.	.
Program 1	Amberley vs Exposed	0.59	0.40	0.88	2	6.99	0.0303
	Richmond vs Exposed	0.67	0.46	0.98	.	.	.
Program 2	Amberley vs Exposed	0.48	0.31	0.75	2	10.72	0.0047
	Richmond vs Exposed	0.57	0.37	0.87	.	.	.
Dose	Mild exposure vs Unexposed	1.49	0.99	2.25	3	9.25	0.0261
	Moderate exposure vs Unexposed	1.45	0.96	2.16	.	.	.
	Prolonged exposure vs Unexposed	1.68	1.13	2.48	.	.	.

In the primary multiple logistic regression analyses (Table 12.15), the odds of having pigmented or sun-related lesions were statistically significantly lower for both the Amberley comparison (0.56; 95% CI: 0.40, 0.78) and the Richmond comparison (0.64; 95% CI: 0.47, 0.87). This relationship was consistent in the Program 1 and 2 subgroup analyses, in the reduced model, and with the robust standard error estimates, although there was no linear trend of increasing odds with increasing dose.

12.5.2 Breast examination

Self-Reported Breast Abnormalities (Postal Questionnaire)

Responses to the four individual Postal Questionnaire items on self-reported breast abnormalities and the combined outcome of any self-reported breast abnormality are presented in Table 12.16. Of the 21 female respondents to the Postal Questionnaire, none reported any breast symptoms or problems. There were too few events to proceed with multivariate analysis, although there was no difference between groups ($\chi^2_2 = 2.9470, p = 0.23$).

Table 12.16 : Number and percentage of participants with self-reported breast abnormalities by group

Self-reported abnormalities	Amberley		Richmond		Exposed	
	N=483	%	N=585	%	N=635	%
Any new lump(s) in the breast area	2	4.0	9	1.5	4	0.6
	N=482	%	N=584	%	N=635	%
Any change to the skin of nipple/breast	2	0.4	3	0.5	7	1.1
	N=482	%	N=584	%	N=635	%
Unusual increase in the size of one breast	1	0.2	3	0.5	2	0.3
	N=482	%	N=585	%	N=635	%
Sticky/bloody discharge-one/both nipples	0	0.0	1	0.2	1	0.2
	N=483	%	N=585	%	N=635	%
Any breast abnormality	3	0.6	11	1.9	12	1.9

Diagnosed Breast Abnormalities (health examination)

There were 12 individuals who did not participate in the breast examination for gynaecomastia or other breast abnormality. Of these, three were female and nine were male. Of the 18 females who agreed to a breast examination, none was identified as having other breast abnormality.

Gynaecomastia

There was no heterogeneity across HSA centres for gynaecomastia in the Richmond versus exposed comparison (p=0.62), but there was substantial heterogeneity for Amberley versus exposed (p=0.004). Given the wide variation in diagnosis of gynaecomastia across centres and the low frequency of gynaecomastia, the data were judged too unreliable to proceed with analysis.

Other breast abnormality

A number of other breast abnormalities apart from gynaecomastia were detected, mainly various breast nodules, lumps or asymmetry. There were 38 breast abnormalities detected in the right breast and 46 in the left breast. When combined, there was a total of 59 participants with a breast abnormality. There was no heterogeneity across HSA centres (p=0.46 for Amberley versus exposed and p=0.15 for Richmond versus exposed). Table 12.17 indicates the distribution of these abnormalities across the three groups. Given that the incidences are all below 5%, there was insufficient statistical power to proceed to regression analyses.

Table 12.17 : Number and percentage of participants with other physician diagnosed breast abnormality by exposure group

	Amberley N=405		Richmond N=513		Exposed N=606	
	n	%	n	%	n	%
Other physician diagnosed breast abnormality	15	3.7	24	4.7	20	3.3

12.6 Discussion

12.6.1 Skin examination

There was some interest *a priori* in lipomata, as they appeared to have been commonly reported in those exposed participants enrolled in the health care scheme. However, because of the heterogeneity of diagnosis of lipomata across HSA examination centres, a combined measure of association between exposure group and this outcome could not be obtained.

There was no detectable association between DSRS exposure and psoriasis, measured either as self-reported physician diagnosis or judged objectively by a physician at the physical examination, in the primary or any of the secondary analyses. This is consistent with current thinking that psoriasis is an immunologic disease of T cell regulation with strong genetic influences and not related to environment.

There appeared to be an increased risk of physician-diagnosed dermatitis in the DSRS exposed group. This association was statistically significant in the comparison with Amberley (OR=1.6; 95% CI: 1.04, 2.4) and on the borderline of significance with the Richmond (OR=1.42; 95% CI: 0.98-2.1). The point estimates were similar in the Program 1 and 2 subgroups, although only the Program 2 comparison with the Amberley group formally reached statistical significance. The dose-response effect was also highly suggestive, with the odds ratios increasing from 1.1 to 1.5 to 1.6 with increasing dose, although these did not reach statistical significance.

This association was also borne out in the analysis of self-reported diagnosis of dermatitis. The odds ratio for self-reported dermatitis was 3.6 (95% CI: 2.5, 5.0) versus the Amberley comparison group and 2.1 (95% CI: 1.6, 2.8) versus the Richmond comparison group. This was also statistically significant for Programs 1 and 2, and, although the point estimates were not completely suggestive of a dose-response effect (2.2 to 3.1 to 2.7 with increasing exposure), all of these were statistically significantly increased. The consistency of the data, and the similarity between the self-reported and physician-diagnosed analyses, lead us to conclude that there is likely to be an adverse effect of DSRS on dermatitis, increasing the risk anywhere between 30% to 200%.

There was a clear association between self-reported skin rash/irritation and DSRS exposures, with an OR=2.6 (95% CI: 2.0, 3.3) versus Amberley and an OR=2.3 (95% CI: 1.8, 2.9) versus Richmond. This remained significant in the subgroup analyses of Program 1 and 2, and there was indication of a dose-response effect, with ORs increasing from 2.1 to 2.4 to 2.8, all of which were statistically significant. Given that this symptom is likely to be related to allergic dermatitis (reported above), this further bolsters the possibility of an adverse effect of DSRS exposures on dermatitis. There was no convincing pattern of association between DSRS exposures and self-reported skin ulcers or self-reported physician diagnosis of eczema.

For physician-diagnosed squamous cell carcinoma, basal cell carcinoma, and malignant melanoma at the time of the examination, there was no evidence of any association with DSRS exposures, and no dose-response effect. However, for self-reported physician diagnosis of malignant melanoma, there appeared to be an increased effect of DSRS compared to the Amberley group (OR=2.0, 95% CI: 1.04, 3.8), but not compared to the Richmond group. This pattern was mirrored in the subgroup analyses of Program 2.

Pigmented or sun-related lesions included all moles, naevi, actinic or solar keratoses, suspicious lesions, etc. that were recorded by physicians; this end point is evidently somewhat subjective and “soft”, although there should not be any differential bias, given that physicians were blinded to the exposure status of the participants during the examination. There was a strong, statistically significant effect of DSRS exposures on these lesions, with the odds ratio being 1.96 (95% CI: 1.3, 2.9) versus Amberley comparisons and 2.1 (95% CI: 1.45, 3.13) versus Richmond comparisons. This was consistent and statistically significant for both comparisons groups in both Program 1 and Program 2 sub-analyses. The point estimates did not clearly indicate a graded dose-response curve, although the confidence intervals overlapped each other. The point estimates for the dose-response analysis were consistently elevated, with the moderate dose group reaching statistical significance.

12.6.2 Breast examination

There was substantial and significant variation in the diagnosis of gynaecomastia between centres, which was *not* due to the variation in participant groups. A test of heterogeneity was strongly significant, indicating that, despite physician training and efforts at standardisation, there was great variability in the diagnosis. Judging gynaecomastia is notoriously difficult since it involves differentiating mammary tissue from subcutaneous fat; and this can be

particularly difficult in more obese men. This variability remained, however, even after adjusting for BMI, indicating that physicians were judging gynaecomastia very differently. It did not make sense to pursue analyses since these results cast great doubt on the very validity of the diagnosis, and we would not be able to interpret results with any degree of confidence. There were too few other breast abnormalities, such as nodules and lumps, to analyse.

12.7 Conclusions

There was a strong and statistically significant two- to three-fold increase in dermatitis in the F-111 DSRS group, and this was consistent between the three methods of assessment: as self-report, previous physician diagnosis, and as observed in the health examination. This effect was more marked versus the Amberley comparisons than the Richmond group. There was a less robust two-fold increase in pigmented or sun-related lesions (i.e. moles, naevi and solar/actinic keratoses) in the DSRS group versus both comparison groups. Other outcomes were either too rare or too variable to be analysed or they showed no difference between groups.

12.8 References

1. Cone, J.E. Health hazards of solvents. *State of the Art Reviews - Occupational Medicine* **1**, 69-87 (1986).
2. Sullivan, J.B., Levine, R.J., Bangert, J.L., Maibach, H. & Hewitt, P. Chapter 15: Clinical Dermatotoxicology. in *Clinical Environmental Health and Toxic Exposures* (ed. Krieger, G.R.) 182-206 (Lippincott Williams & Wilkins, Philadelphia, 2001).
3. Ritchie, G. et al. Biological and health effects of exposure to kerosene-based jet fuels and performance additives. *Journal Toxicology and Environmental Health, Part B Critical Reviews* **6**, 357-451 (2003).
4. Kaufman, J.D., Cohen, M.A., Sama, S.R., Shields, J.W. & Kalat, J. Occupational skin diseases in Washington State, 1989 through 1993: using workers' compensation data to identify cutaneous hazards. *American Journal of Public Health* **88**, 1047-51 (1998).
5. Hackett, J.P. Allergic contact dermatitis in American aircraft manufacture. *American Journal of Contact Dermatitis* **10**, 157-66 (1999).

6. Bock, M., Schmidt, A., Bruckner, T. & Diepgen, T.L. Occupational skin disease in the construction industry. *British Journal of Dermatology* **149**, 1165-71 (2003).
7. Chen, R., Semple, S., Dick, F. & Seaton, A. Nasal, eye, and skin irritation in dockyard painters. *Occupational & Environmental Medicine* **58**, 542-543 (2001).
8. Kanerva, L., Jolanki, R., Estlander, T. & Alanko, K. Latent (subclinical) contact dermatitis evolving into occupational allergic contact dermatitis from extremely small amounts of epoxy resin. *Contact Dermatitis* **43**, 47-9 (2000).
9. Banney, L.A., Weedon, D. & Muir, J.B. Nuchal fibroma associated with scleredema, diabetes mellitus and organic solvent exposure. *Australasian Journal of Dermatology* **41**, 39-41 (2000).
10. Rugo, H.S. & Fischman, M.L. Occupational Cancer. in *Occupational & Environmental Medicine* (ed. LaDou, J.) 235-272 (Lange Medical Books/McGraw-Hill, New York, 1997).
11. Labreche, F.P. & Goldberg, M.S. Exposure to organic solvents and breast cancer in women: a hypothesis. *American Journal of Industrial Medicine* **32**, 1-14 (1997).
12. Blair, A., Hartge, P., Stewart, P.A., McAdams, M. & Lubin, J. Mortality and cancer incidence of aircraft maintenance workers exposed to trichloroethylene and other organic solvents and chemicals: extended follow up. *Occupational and Environmental Medicine* **55**, 161-171 (1998).
13. Mabuchi, K., Bross, D.S. & Kessler, I.I. Risk factors for male breast cancer. *Journal of the National Cancer Institute* **74**, 371-375 (1985).
14. Hansen, J. Breast cancer risk among relatively young women employed in solvent-using industries. *American Journal of Industrial Medicine* **36**, 43-47 (1999).
15. Steenland, K., Stayner, L. & Deddens, J. Mortality analyses in a cohort of 18 235 ethylene oxide exposed workers: follow up extended from 1987 to 1998. *Occupational and Environmental Medicine* **61**, 2-7 (2004).

13 Neurological Outcomes

Chapter summary

Neurological outcomes assessed in this study included: a) colour vision, measured in two ways (by the colour confusion index, which is a number; and as a clinical diagnosis, i.e. blue-yellow, red-green, mixed or non-specific deficits); b) peripheral vibration sensation in the hands and feet (using a biothesiometer [a calibrated vibrating probe] and self report); and c) olfaction (using the "Sniffin' Sticks" battery, a series of odourant pens). Results indicate a definite two- to three-fold increase in subjective sensory and motor neuropathic symptoms in the DSRS group relative to both comparison groups, but this is not supported to any degree by the objective vibration sense tests. There was a consistent increase of approximately 35% in impaired colour vision in the exposed group versus the Richmond comparisons. There was no detectable change in olfaction objectively, although there was a three-fold increase in self-reported sensitivity to smells.

Chapter contents

13.1 Introduction

The impetus for including neurological outcomes in SHOAMP was partly from the Board of Inquiry (BOI) findings, partly from the literature review, and partly from presentations to the F-111 Interim Health Care Scheme. The latter registered complaints including motor neurone disease, multiple sclerosis, blue-yellow colour vision loss, and loss of feeling in fingers and toes. A number of neurological outcomes were targeted for study, and these are detailed below along with their rationale.

Well-known occupational neurotoxins that affect the peripheral nervous system include the solvents *n*-hexane and methyl-*n*-butyl-ketone; the metals lead, mercury, arsenic, and thallium; the organophosphate pesticides; and the gases carbon disulfide and ethylene oxide. Exposures relevant to DSRS in this regard include jet fuel,[1] organic solvent mixtures[2-5] and solvents associated with painting.[6-8] While inherited colour vision deficiencies are almost entirely in the red-green range,[9-11] acquired deficiencies are usually in the blue-yellow range. In addition, in acquired dyschromatopsia* the deficiencies can involve the eyes unequally or they can be monocular, and they can have a variable, progressive or regressive course, depending on various factors.[12-15] Of particular interest to the SHOAMP study are the studies which report that visual evoked potentials, vision, and visual contrast sensitivity, are affected by exposure to *n*-hexane, toluene, ethyl-benzene and solvent mixtures. Table 13.1 provides a summary of the main relevant exposure substances known to affect colour vision that are relevant to DSRS.

Olfactory neurons are the only central nervous system cells directly in contact with the environment and are therefore potentially harbingers of neurotoxic effects.[35] Such damage can lead to a permanent impairment of a sense-organ. Olfactory dysfunction has been reported after high exposure to hydrogen sulphide cadmium, methyl mercury, toluene diisocyanate, manganese, sulfur dioxide, xylene acrylates, methacrylates, ammonia, and organic solvent compounds.[36-41] The substances relevant to DSRS are detailed in Table 13.2.

* Aquired Dyschromatopsia is a progressive loss of colour vision due to retinal diseases.

Table 13.1 : Potential DSRS exposures shown previously to affect colour vision

Exposure	Study Reference
***n*-hexane**	Raitta et al., 1978[16], Chang, 1987[17]
Toluene	Cavalleri et al., 2000[18], Muttray et al, 1999[19], Urban & Lukas, 1990[20], Vcra et al., 1997[21], Zavalic etal, 1998[22], Zavalic et al, 1998[23], Zavalic et al, 1998[24], Zavalic et al, 1996[25]
Solvent mixtures	Blain & Mergler, 1986[26], Broadwell et al., 1995[5], Donoghue et al., 1995[27], Gonzalez et al., 1998[28], Mergler et al., 1988[29], Mergler & Blain, 1987[13], Muttray et al., 1997[30], Nolfe et al., 1991[31], Pratt et al., 2000[32], Semple et al., 2000[33]
Ethyl benzene	Druzdz et al., 1997[34]

Table 13.2 : Potential DSRS exposures associated with olfactory dysfunction in humans

Substances	Substances
Acetic acid	Organic fluorine compounds
Acetone	Xylene
Benzene	Aluminium fumes
Butyl acetate	Chromium compounds
Ethyl acetate	Silica
Ethyl ether	Strontium sulphate
Halogenated organic compounds	Isocyanates

Adapted from: Sullivan, J. B. in Clinical Environmental Health and Toxic Exposures (eds. Sullivan, J. B. & Krieger, G. R.) 390-396 (Lippincott Williams & Wilkins, Philadelphia, 2001).

Motor Neurone Disease (MND) is a progressive affliction of the central nervous system, characterised by a degeneration of spinal and/or bulbar motor neurones and corticospinal tracts. It usually develops in late adult life and affects males more commonly than females. MND has an invariably progressive and fatal course. Whilst a hereditary form of MND is well recognised, environmental and/or occupational chemical exposures have been associated with MND. Various exposures have been suggested, including work in the leather industry,[42,43] organic solvents,[44] and particular jobs including farming, professional contacts with animal hides and carcasses, working textile factories,[45] firefighters, military personnel and janitors.[46]

Multiple Sclerosis (MS) is the result of damage to myelin, the protective sheath surrounding the nerve fibres of the central nervous system. Although the exact cause is unknown, it appears to result from a malfunction of the immune system, which then starts to attack its host body. Recent research has shown that though MS is not hereditary, the presence of certain genes can make an individual susceptible to this malfunction, which may also be triggered by a viral infection.[47] Environmental causes are uncertain, although organic solvents might contribute.[48]

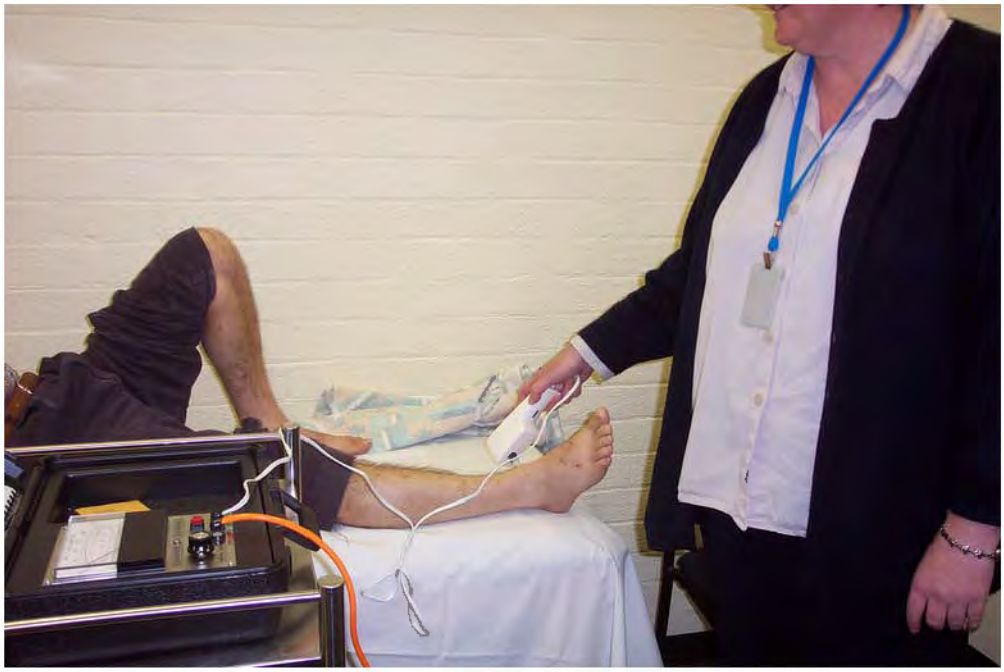
13.2 Measures

13.2.1 Peripheral neuropathy

Several testing procedures can be applied to investigate the effect of occupational exposure on tactile sense, including vibration perception threshold (VPT), two-point discrimination, depth sense perception, temperature threshold, pain threshold and other sensations. To date, VPT is the most commonly adopted method for occupational epidemiological studies in groups of workers, as it has the property of detecting early change.[37] Several studies have used VPT measurement for early detection of chemical-related neuropathies.[2,4,49,50]

For SHOAMP, VPT was measured for the participants in the general health and medical examination using a biothesiometer (Bio-Medical Instrument Co., Newbury, OH, USA). This product is an efficient tool for the screening of workers with significant exposure to neurotoxins or with early sensory symptoms.[51] Participants were asked to lie on their back on a bed while the attending HSA doctor conducted this examination (Figure 13.1).

Figure 13.1 : Vibration sensation testing



The VPT was measured by increasing the amplitude of the vibrator button from zero to the point at which the vibration is just perceptible. With the vibrator button applied to the point to be tested, the participant was asked to concentrate their attention on the test and to report the first appearance of the sensation of vibration by saying “now”. In each anatomical location, the VPT was recorded as the lowest voltage at which vibration could be sensed. Lower and upper limbs were assessed in turn (both right and left sides), with voltage being increased by 1 millivolt every 2 to 3 seconds until the participant became aware of vibration. The threshold was measured twice in the following locations:

- dorsal surface of the big toe
- medial malleolus (ankle)
- medial side of knee (tibial plateau)
- distal interphalangeal joint of middle finger
- metacarpophalangeal joint (hand)
- radial styloid (wrist)
- olecranon (elbow).

13.2.1.1 Self-reported symptoms of neurological disorders

In addition to the physical testing, self-reported neurological signs/symptoms were also recorded by participants in their Postal Questionnaire. The participants were asked the series of yes/no questions outlined below.

Sensory symptoms

2.68 In the past month have you suffered from difficulty recognising hot from cold water?
2.69 In the past month have you suffered from difficulty feeling pain, cuts or injuries?
2.74 In the past month have you suffered from numbness, “asleep feeling” or prickling sensation in hands or arms?
2.75 In the past month have you suffered from numbness, “asleep feeling” or prickling sensation in feet or legs?
2.78 In the past month have you suffered from unusual sensitivity or tenderness of your skin when clothes or bedclothes rub against you?

For the purpose of analysis these measures were combined, and a participant was considered to have sensory symptoms if they answered yes to any of these questions.

Motor symptoms

2.65 In the past month have you suffered from difficulty undoing buttons?
2.68 In the past month have you suffered from problems with tripping, or your feet slapping, while walking?
2.70 In the past month have you suffered from feeling unsteady walking on uneven ground?
2.71 In the past month have you suffered from feeling unsteady walking in the dark?

For the purpose of analysis these measures were combined, and a participant was considered to have motor symptoms if they answered yes to any of these questions.

13.2.2 Colour vision

Although both the L’Anthony and the Fairnsworth D-15 tests for colour vision were administered, the L’Anthony is the desaturated version of the test and detects more subtle colour vision defects; hence only this measure is considered for this report. These tests consist of placing a series of 15 coloured caps (each with a hidden number) in order of

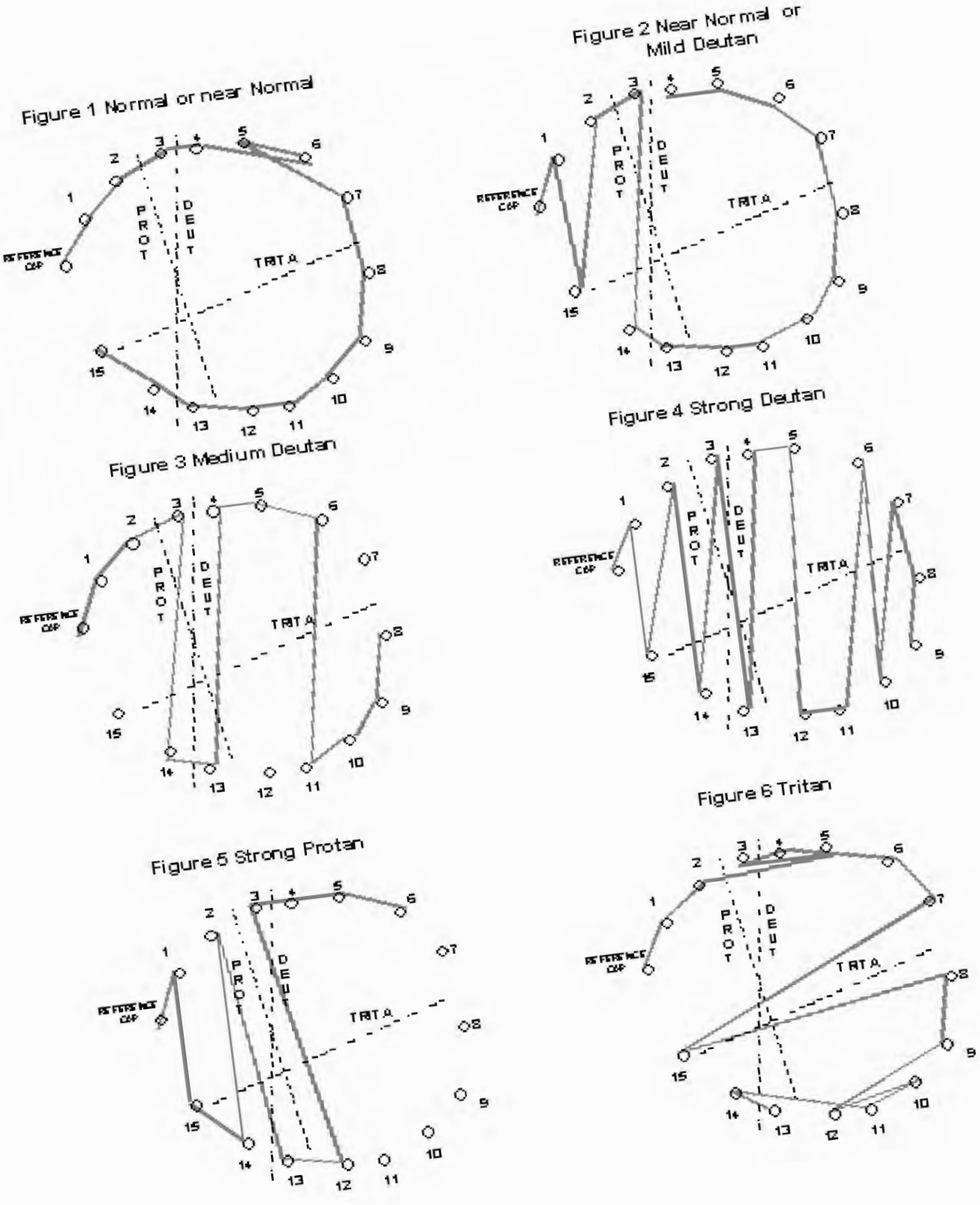
gradually-changing tint. In scoring the test, a specifically-designed pre-printed circular chart was used; the numbers represent the correctly-ordered colour caps consecutively in clockwise order around this circle or wheel (see Figure 13.2).

Starting from the reference cap, a line is drawn linking caps in the order in which they are placed by the individual. For individuals with perfect colour vision, the resulting figure is an incomplete circle with a line connecting each consecutive number (i.e. the connecting lines are located around the circumference of this circle or wheel). For individuals with less than perfect colour vision, the order of the caps is not consecutively located around the circumference of the wheel. The lines connecting the numbers are not located on the circumference of the circle, but rather the lines may “jump” from one part of the circle to the other. These lines which do not connect consecutive numbers on the circle are called “crossings”. It is not uncommon for individuals to have small or minor crossings, for example interchanging two consecutive numbers. This is demonstrated in the first chart in Figure 13.2, where an individual has ordered the caps as …, 4, 6, 5, 7, … rather than the correct ordering of .., 4, 5, 6, 7, … A major crossing occurs when the line crosses from one side of the circle to another, as demonstrated in diagrams 2–6 (in Figure 13.2).

Colour vision deficits are demonstrated not only by the number of crossings, but also by the axis (or angle) on which the crossing occurs. The axis of a crossing is referred to as the “confusion axis”, and it is the confusion axis which determines the type of deficit, of which there are three: deutan, protan and tritan.

Deutan is the most common type of vision deficiency and it affects mainly the green receptors. An individual with deutan loss will have trouble distinguishing blue-green from grey and red-purple. Deutan subjects exhibit a “confusion axis” from green to purple. Protan is a vision deficiency affecting mainly the red receptors. An individual with protan loss will have trouble distinguishing red-green and will confuse red-orange with blue-green and grey. Protans have a “confusion axis” from red to blue-green. Tritan affects mainly the blue receptors. An individual with a tritan deficit will confuse violet with grey and yellow-green. Tritan loss is rarely inherited. It shows a “confusion axis” from yellow to blue.

Figure 13.2 : L'Anthony descriptors of vision loss



13.2.3 Olfaction

Investigating olfactory function in occupationally exposed workers is problematic for various reasons[37]. The main difficulty is the lack of agreement on testing procedures. Odour detection threshold, discrimination, or identification are the most frequently used measures for studies in groups of workers. Commercially available methods such as the University of Pennsylvania Identification Test (UPSIT)[52] or the Connecticut Chemosensory Clinical Research Centre Test (CCCRCT)[53] are frequently applied in studies performed in the US, whereas in Europe the "Sniffin' Sticks" test battery, based on the Kobal and Hummel method,[54] or Elsberg and Levy's blast injection method modified by Pruszewic[55], or a combination of different methods, were most commonly adopted. The "Sniffin' Sticks" test battery was selected for SHOAMP because a research group based at Queensland's Griffith University had previously collected Australian normative data using it.

The "Sniffin' Sticks"[56] use pen-like odour-dispensing devices presented by an HSA doctor to the participant. Olfaction assessment comprises three tests (called the Extended Identification Test):

- odour threshold (*n*-butanol, testing by means of a single staircase)
- odour discrimination (16 pairs of odourants, triple forced choice)
- odour identification (16 common odourants, multiple forced choice from four verbal items per test odourant).

SHOAMP procedure

The attending HSA doctor was instructed to wear odourless gloves. Perfume and strongly-scented deodorant also had to be avoided. For odour presentation, opened pens were positioned approximately 2 cm in front of both of the participant's nostrils. The participant was then asked to sniff while each pen was presented for no longer than 3-4 seconds each. When lateralised tests were performed, pens were presented to the left or right nostril; the patient closed the contralateral nostril by bringing the soft side of the thumb to the naris. Testing was performed in a quiet, air-conditioned room, with the patients receiving no ongoing feedback as to the quality of their performance.

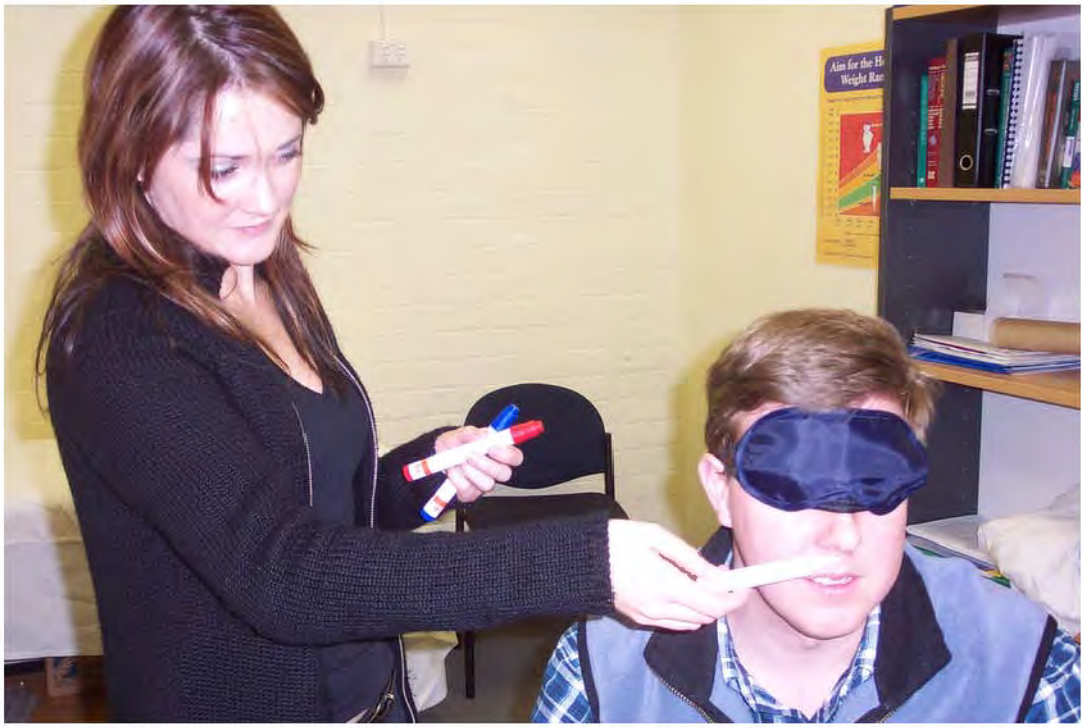
Fifteen minutes before testing commenced, participants were not permitted to eat or drink anything but water and were required to stay away from chewing gum, sweets and cigarettes. The threshold component of the olfaction test was performed first, followed by the discrimination component, and then the identification component.

Threshold test (numbers on pens in red)

Blindfolded participants were presented with triplets of odourants, with each pen from a triplet presented at intervals of five seconds each. One of the pens contained the odourant (red cap), the others were almost odourless (green and blue caps). Participants had to identify the pen that smelt different from the other two pens (Figure 13.3).

There was a 30-second interval between presentation of the first pen of a triplet and presentation of the first pen of the following triplet, with each pen presented only once. The first triplet was presented in the sequence RED, GREEN, BLUE. Next was BLUE, RED, GREEN. Next, GREEN, BLUE, RED. This cycle had to be repeated throughout the entire procedure.

Figure13.3 : Olfaction testing



Pens were presented in ascending concentration until the participant correctly identified the odour. A concentration step was correctly identified when the odour-containing pen had been identified correctly twice in a row. Triplets were only presented twice when the odour-containing pen had been correctly identified on the first presentation of this concentration step. As soon as the starting point (which was marked on the Threshold record sheet) had been identified a one-step lower concentration was presented (with a higher number

indicating a higher dilution step). This was continued until the participant missed one concentration. This miss triggered a reversal of the staircase. If this concentration was missed again, the next (higher) concentration was administered until a dilution step had been correctly identified. Testing was finished when seven reversals of the staircase had been found (including the starting point).

The result sheet for the Threshold test had seven columns, where a single mark was recorded in each. The number of that mark corresponded to the number of the row. The score was the average of the last four columns. For example, in the last column, row 6 was marked; in the second-last column, row 7 was marked; in the third-last column, row 6 was marked; and in the fourth-last column, row 8 was marked. The score is (6+7+6+8) divided by 4, out of a total of 16.

Discrimination test (numbers on pens in green)

Sixteen triplets were presented to the blindfolded participant; two pens had the same smell, with one of the three pens containing a different smell. The participant had to identify the pen that smelt different. There was a 30-second interval between presentation of the first pen of a triplet and presentation of the first pen of the following triplet, with each pen presented only once. The sequence of the three pens had to be changed by the investigator. The first triplet was presented as RED, GREEN, BLUE. Next, BLUE, RED, GREEN. Then GREEN, BLUE, RED. The row GREEN had the correct answers, so the score reflects the total of the answers in row GREEN, out of a total of 16.

Identification test (numbers on pens in blue)

Blindfolded, each participant was presented with 16 pens at intervals of 30-seconds. Each person was given a sheet of paper which described 16 lists of four items each. Participants had to identify the item from each list that best described the presented odour. One point per correct answer was given, with a maximum score of 16. The total olfaction score calculated from all sub-test results for each participant was called the TDI (Threshold, Discrimination, Identification) and represented the sum of the results for all three tests.

Before taking the olfaction test, participants were asked if they currently suffered from epilepsy, upper respiratory tract disease, influenza, acute asthma, hay fever or the common cold. Administration of the olfaction test could potentially exacerbate epilepsy and acute asthma, therefore participants who reported currently having those conditions were considered ineligible for assessment. The presence of upper respiratory tract disease,

influenza, hay fever or a cold could interfere with the ability to smell. The olfaction test was not administered to patients reporting any of these conditions.

Two additional screening questions required identification of any other relevant conditions which could interfere with the test. Of particular interest were Alzheimer's disease and Parkinson's disease, which are known to affect the ability to recognise odours. Participants who reported Alzheimer's Disease or Parkinson's Disease were excluded from analysis.

Further, one question in the Postal Questionnaire (item 2.46) asked if the participant had any increased sensitivity to odours, and this was included as an additional outcome.

13.2.4 Diagnosed or treated medical conditions

The Postal Questionnaire asked whether the participant had ever been diagnosed with, or treated for, the following medical conditions:

2.86 Motor Neurone Disease (MND)

2.87 Multiple Sclerosis (MS).

13.3 Potential confounders

The following potential confounders were considered for peripheral neuropathy: age, posting category, rank category, alcohol consumption, smoking status, diabetes, height, and civilian occupational exposures to organic solvents and the neurotoxin lead. For outcomes measured as part of the health examination, HSA centre was also included.

In addition to the general potential confounders (age, posting and rank) described in Chapter 8, visual acuity and diabetes were additional potential confounders in analysing colour vision.

Potential confounders of interest for the analysis of olfaction data were age, exposure or posting category, rank, smoking status (which is known to affect smell), civilian solvent exposure, and HSA centre. Although more than 60 medications have been known to alter smell[58], this is a rare adverse effect, restricted mainly to highly-specialised drugs, and unlikely to be a large confounder in this study.

13.4 Analyses

All analyses followed the outline described in Chapter 6 (Analysis).

13.4.1 Peripheral neuropathy

Only the vibration perception threshold measures from the most distal site on the fingers and toes were analysed because these are the most sensitive areas. The results of VPT were analysed in two ways:

1) *Quantitatively* – as measured directly with the biothesiometer. The initial scores range from 0 to 50. The repeated (two) measures conducted at each anatomical location were averaged for each side (left and right). The highest (worst) average value between left and right side was then taken as the measure for that site.

2) *Clinical categories*. A clinical assessment of each participant was undertaken by categorising VPT into diagnosis categories of normal, slight impairment, moderate, severe, and extreme. The categories were formed using the guidelines established by Cornblath for the vibration sensation component of the Total Neuropathy Score.[59] The categories were:
   i) Normal sensation – normal to 125% above normal
   ii) Slightly reduced sensation – 126% to 150% above normal
   iii) Moderately reduced sensation – 151% to 200% above normal
   iv) Severely reduced sensation – 201% to 300% above normal
   v) Extremely reduced sensation – > 300% above normal.

Frequencies were obtained for each of these categories, and the categories were then combined into normal (category 1) or not normal (categories 2-5) for regression analyses.

13.4.2 Colour vision

The results of the L'Anthony D15-D were scored in two ways:

1) *Quantitatively* – as Bowman's Colour Confusion Index (CCI). This was calculated as the ratio of the sum of the colour differences between adjacent caps as positioned by

the participant, relative to the sum of the differences between caps correctly positioned. For a perfect score this ratio is 1. The more mistakes or “crossings” made by the participant, the greater the distance between consecutively-placed caps and the higher the CCI. For SHOAMP, colour vision data are presented as the participant’s highest CCI (worst eye) of the two eyes tested.

2) *Qualitatively* – as a clinical diagnosis. A clinical assessment of each participant was undertaken by plotting lines between consecutive caps on a vision assessment chart (as seen in Figure 8.1). Based on the major confusion axis, a diagnosis of normal, red-green, blue-yellow, mixed or non-specific was made for each eye. These results were subsequently categorised as normal, not-normal.

Because of the extremely right-skewed distribution of CCI scores, and to aid clinical interpretation of the results, the CCI was categorised into three groups: normal, slight abnormality, and moderate/severe abnormality. A normal classification was assigned if the participant had a perfect CCI score or a score less than the highest CCI score achieved by an individual with a normal clinical assessment (CCI of < 1.29). The median of the remaining CCI scores was used to define the mildly abnormal group (i.e. those less than the median) and the moderately or severely abnormal group (i.e. those equal to or greater than the median); the latter two were collapsed for analysis.

13.4.3 Olfaction

While participants who reported the presence of any of the six screening conditions were supposed to be excluded from attempting the olfaction test, this did not always occur in practice. Irrespective of whether or not they undertook the olfaction test, any participant reporting the presence of current upper respiratory tract disease, influenza, hay fever or the common cold was excluded from analyses. Any participant who reported the presence of epilepsy or acute asthma but who completed the olfaction test was included in analyses, as these conditions do not affect the validity of the measure. Participants were also excluded from analyses if the clinician reported Alzheimer’s Disease or Parkinson’s Disease as an “other” or “neurological” condition.

The total olfaction score was obtained by summing the score for each of the three individual components: threshold, discrimination and identification. The range of scores was from 0 to 48 (0 to 16 for each component), with a higher score indicating a better sense of smell.

Participants needed to complete the last four steps of the threshold component and all 16 items for each of the discrimination and identification components of the olfaction test in order to be scored. The outcome of interest (i.e. total olfaction score) was analysed as a continuous variable. Further, one question in the Postal Questionnaire (item 2.46) asked if the participant had any increased sensitivity to smells or odours in the past month.

13.5 Results

13.5.1 Peripheral neuropathy

A total of 1538 SHOAMP participants received a health examination and 1726 participants returned a completed Postal Questionnaire. Four participants did not undertake VPT testing and were thus excluded from the peripheral neuropathy analysis. Twenty-two participants did not complete the questions relating to neurological self-reported symptoms and were excluded from the self-reported neurological symptoms analysis. Forty-two individuals did not complete the question on Motor Neurone Disease (MND) and 40 did not complete the question on Multiple Sclerosis (MS).

13.5.1.1 Vibration Perception Threshold score

The distribution of the VPT scores was found to be skewed to the right, as a large number of participants had relatively normal scores. A normal distribution was obtained with a log transformation. As seen in Table 13.3, the median and inter-quartile range of log VPT scores are similar across exposure groups for both finger and great toe. There was no evidence of heterogeneity across HSA centres ($F = 1.23$, $p = 0.25$ for finger and $F = 0.65$, $p = 0.82$ for toe). See Table 13.4 for the multiple regression analyses.

Table 13.3 : Log distribution of VPT scores for finger and toe

Measure of VPT	Amberley N = 405		Richmond N = 516		Exposed N = 613	
	Median	Quartiles	Median	Quartiles	Median	Quartiles
Finger	1.6	1.4 1.9	1.6	1.4 1.9	1.6	1.4 1.9
Big toe	1.6	1.4 1.9	1.6	1.4 1.9	1.6	1.4 1.9

Table 13.4 : Log vibration perception threshold score of the finger – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Estimate	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	-0.04	-0.09	0.01	0.20	2	0.1027
	Richmond vs Exposed	-0.05	-0.09	-0.00	.	.	.
Program 1	Amberley vs Exposed	-0.03	-0.09	0.03	0.20	2	0.5053
	Richmond vs Exposed	-0.03	-0.09	0.03	.	.	.
Program 2	Amberley vs Exposed	-0.05	-0.11	0.02	0.20	2	0.1997
	Richmond vs Exposed	-0.06	-0.12	0.01	.	.	.
Dose	Mild exposure vs Unexposed	0.03	-0.04	0.09	0.19	3	0.0738
	Moderate exposure vs Unexposed	0.08	0.02	0.14	.	.	.
	Prolonged exposure vs Unexposed	0.02	-0.04	0.08	.	.	.

The multiple regression analyses showed that the log VPT scores at the finger were slightly lower (better) for the Amberley and Richmond comparison groups, relative to the exposed group (Table 13.4). However, this difference was only significant for the Richmond comparison group in the primary analyses, with a difference of –0.05 log units (95% CI: -0.09, -0.00). This remained significant in the reduced model, and both the Amberley and Richmond comparisons were significant with the robust standard error estimates. Program 1 and 2 subgroup analyses were not significant (p=0.51 and p=0.20 respectively) and there was no clear dose-response effect. Multiple regression results for the log VPT scores at the toe (Table 13.5) indicated no effect overall (p=0.72), in Program 1 (p=0.54), in Program 2 (p=0.74), and there was no dose-response effect. This lack of significance remained in the reduced model and with the robust standard error estimates.

Table 13.5 : Log vibration perception threshold score of the toe – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Estimate	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	-0.01	-0.06	0.05	0.39	2	0.7166
	Richmond vs Exposed	-0.02	-0.08	0.03	.	.	.
Program 1	Amberley vs Exposed	0.04	-0.03	0.11	0.40	2	0.5441
	Richmond vs Exposed	0.03	-0.04	0.09	.	.	.
Program 2	Amberley vs Exposed	-0.01	-0.08	0.06	0.41	2	0.7379
	Richmond vs Exposed	-0.03	-0.10	0.05	.	.	.
Dose	Mild exposure vs Unexposed	0.02	-0.06	0.09	0.39	3	0.5514
	Moderate exposure vs Unexposed	0.04	-0.03	0.11	.	.	.
	Prolonged exposure vs Unexposed	-0.02	-0.09	0.05	.	.	.

13.5.1.2 Vibration Perception Threshold: clinical assessment

Table 13.6 and Table 13.7 show the distribution for clinical categories across exposure groups for finger and great toe sites respectively.

Table 13.6 : Number and percentage of participants by clinical category for finger by exposure group

Level of sensation	Amberley N = 405		Richmond N = 516		Exposed N = 616	
	n	%	n	%	n	%
Normal	367	*91*	467	*91*	537	*88*
Slightly reduced	20	*4.9*	26	*5.0*	37	*6.0*
Moderately reduced	15	*3.7*	16	*3.1*	31	*5.1*
Severely reduced	2	*0.5*	6	*1.2*	5	*0.8*
Extremely reduced	1	*0.3*	1	*0.2*	3	*0.5*

Table 13.7 : Number and percentage of participants by clinical category for big toe by exposure group

Level of sensation	Amberley N = 405		Richmond N = 516		Exposed N = 616	
	n	%	n	%	n	%
Normal	243	*60*	319	*62*	350	*57*
Slightly reduced	97	*24*	119	*23*	146	*24*
Moderately reduced	33	*8.2*	40	*7.8*	55	*9.0*
Severely reduced	23	*5.7*	19	*3.7*	43	*7.0*
Extremely reduced	9	*2.2*	19	*3.7*	19	*3.1*

The number of participants in these categories appeared similar for the three exposure groups. These categories were collapsed into normal/abnormal for analyses. While there was no heterogeneity across HSA centres for abnormal level of sensation at the toe ($\chi^2_7 = 8.70, p = 0.28$), there was some indication of heterogeneity for loss of sensation at the

finger ($\chi^2_7 = 17.60, p = 0.014$). There were very small numbers and/or zero values in some cells which may effect these analyses, and given that there was no heterogeneity for the continuous score, logistic regression analyses were undertaken on these data. Table 13.8 shows the multivariate regression analysis for the dichotomised diagnosis at the finger. There was no group effect overall (p=0.51), in Program 1 (p=0.25) or in Program 2 (p=0.94), and no dose-response effect (p=0.35).

Table 13.8 : Relationship between exposure and clinical category of the finger – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.84	0.51	1.36	2	1.36	0.5078
	Richmond vs Exposed	0.76	0.47	1.22	.	.	.
Program 1	Amberley vs Exposed	0.72	0.42	1.24	2	2.61	0.2713
	Richmond vs Exposed	0.66	0.39	1.11	.	.	.
Program 2	Amberley vs Exposed	1.13	0.53	2.42	2	0.12	0.9438
	Richmond vs Exposed	1.06	0.50	2.27	.	.	.
Dose	Mild exposure vs Unexposed	1.17	0.63	2.18	3	3.30	0.3483
	Moderate exposure vs Unexposed	1.66	0.95	2.88	.	.	.
	Prolonged exposure vs Unexposed	1.08	0.58	1.99	.	.	.

The multivariate regression analyses for the dichotomised outcome for the toe are shown in Table 13.9. There was no group effect overall (p=0.37), in Program 1 (p=0.48) or Program 2 (p=0.20), and again there was no dose-response effect (p=0.90). For the clinical assessment of VPT both analyses at the finger and toe gave similar results with the reduced model and with the robust standard error estimates.

Table 13.9 : Relationship between exposure and clinical category of the big toe – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.99	0.71	1.39	2	1.99	0.3688
	Richmond vs Exposed	0.81	0.59	1.12	.	.	.
Program 1	Amberley vs Exposed	1.22	0.81	1.84	2	1.47	0.4799
	Richmond vs Exposed	1.01	0.68	1.50	.	.	.
Program 2	Amberley vs Exposed	0.87	0.55	1.35	2	3.18	0.2040
	Richmond vs Exposed	0.69	0.45	1.08	.	.	.
Dose	Mild exposure vs Unexposed	1.14	0.74	1.75	3	0.57	0.9041
	Moderate exposure vs Unexposed	1.13	0.74	1.72	.	.	.
	Prolonged exposure vs Unexposed	1.08	0.71	1.63	.	.	.

13.5.1.3 Self-reported symptoms of peripheral neuropathy

Sensory symptoms

Sensory symptoms were assessed for the 1705 participants who answered the questions. A total of 802 participants had self-reported sensory symptoms in the past month (Table 13.10 shows the number and proportion of participants).

Table 13.10 : Number and percentage of participants with self-reported sensory symptoms by exposure group

	Amberley N = 494		Richmond N = 591		Exposed N = 633	
	n	***%***	**n**	***%***	**n**	***%***
Difficulty recognising hot from cold water	2	*0.4*	3	*0.5*	21	*3.3*
	n	***%***	**n**	***%***	**n**	***%***
Difficulty feeling pain, cuts or injuries	12	*2.5*	12	*2.0*	64	*10*
	N=487	***%***	**N=591**	***%***	**N=632**	***%***
Suffer from numbness or prickling sensation in hands or arms	135	*28*	173	*29*	271	*43*
	N=485	***%***	**N=591**	***%***	**N=631**	***%***
Suffer from numbness or prickling sensation in feet or legs	117	*24*	126	*21*	243	*39*
	N=486	***%***	**N=591**	***%***	**N=632**	***%***
Unusual sensitivity or tenderness of skin	37	*7.6*	34	*5.8*	95	*15*
	N=487	***%***	**N=591**	***%***	**N=631**	***%***
Total self-report of any sensory symptom	195	*40*	237	*40*	370	*59*

The prevalence of self-reported symptoms was much higher in the exposed group relative to both comparison groups. Multiple logistic regression analyses demonstrated a highly statistically significant association between exposure and self-reported sensory symptoms (Table 13.11). The odds ratios were around 0.5 for both Amberley and Richmond comparison groups, relative to all exposed and Program 1 and Program 2 subgroups ($p<0.0001$ for all comparisons). This result was also consistent in the reduced model. A moderate dose-response relationship was also demonstrated, with odds ratios of 1.9, 2.1 and 2.1 for mild, moderate and prolonged exposure to DSRS.

Table 13.11 : Relationship between exposure and self-report of sensory symptoms – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.52	0.39	0.68	2	33.96	<.0001
	Richmond vs Exposed	0.49	0.38	0.64	.	.	.
Program 1	Amberley vs Exposed	0.46	0.32	0.66	2	25.10	<.0001
	Richmond vs Exposed	0.44	0.32	0.62	.	.	.
Program 2	Amberley vs Exposed	0.53	0.36	0.77	2	15.61	0.0004
	Richmond vs Exposed	0.50	0.35	0.71	.	.	.
Dose	Mild exposure vs Unexposed	1.89	1.31	2.72	3	34.28	<.0001
	Moderate exposure vs Unexposed	2.12	1.49	3.03	.	.	.
	Prolonged exposure vs Unexposed	2.13	1.50	3.01	.	.	.

Motor symptoms

Motor symptoms were assessed for the 1705 participants who answered the questions. A total of 400 participants reported motor symptoms in the past month (Table 13.12). The prevalence of self-reported symptoms was much higher in the exposed group relative to both comparison groups.

Table 13.12 : Number and percentage of participants with self-reported motor symptoms by exposure group

	Amberley N = 486		Richmond N = 591		Exposed N = 633	
	n	***%***	**n**	***%***	**n**	***%***
Difficulty undoing buttons	15	*3.1*	22	*3.7*	70	*11*
	N=486	***%***	**N=591**	***%***	**N=631**	***%***
Problems with tripping, or feet slapping while walking	36	*7.4*	43	*7.3*	109	*17*
	N=485	***%***	**N=591**	***%***	**N=633**	***%***
Feeling unsteady walking on uneven ground	54	*11*	39	*6.6*	124	*20*
	N=485	***%***	**N=590**	***%***	**N=632**	***%***
Feeling unsteady walking in the dark	41	*8.5*	50	*8.5*	142	*22*
	N=486	***%***	**N=591**	***%***	**N=632**	***%***
Total self-report of any motor symptom	85	*17*	94	*16*	221	*35*

The odds of self-reported motor symptoms in the past month were statistically significantly lower for both Amberley and Richmond groups relative to the exposed group. The odds ratios for all exposed, Program 1 and Program 2 analyses were between 0.32 and 0.42 (group effect $p<0.0001$ for all analyses) and this was consistent in the reduced model. There was also a possible dose-response relationship, with odds ratios increasing from 2.4, to 2.4 to 3.2 with increasing level of exposure (Table 13.13).

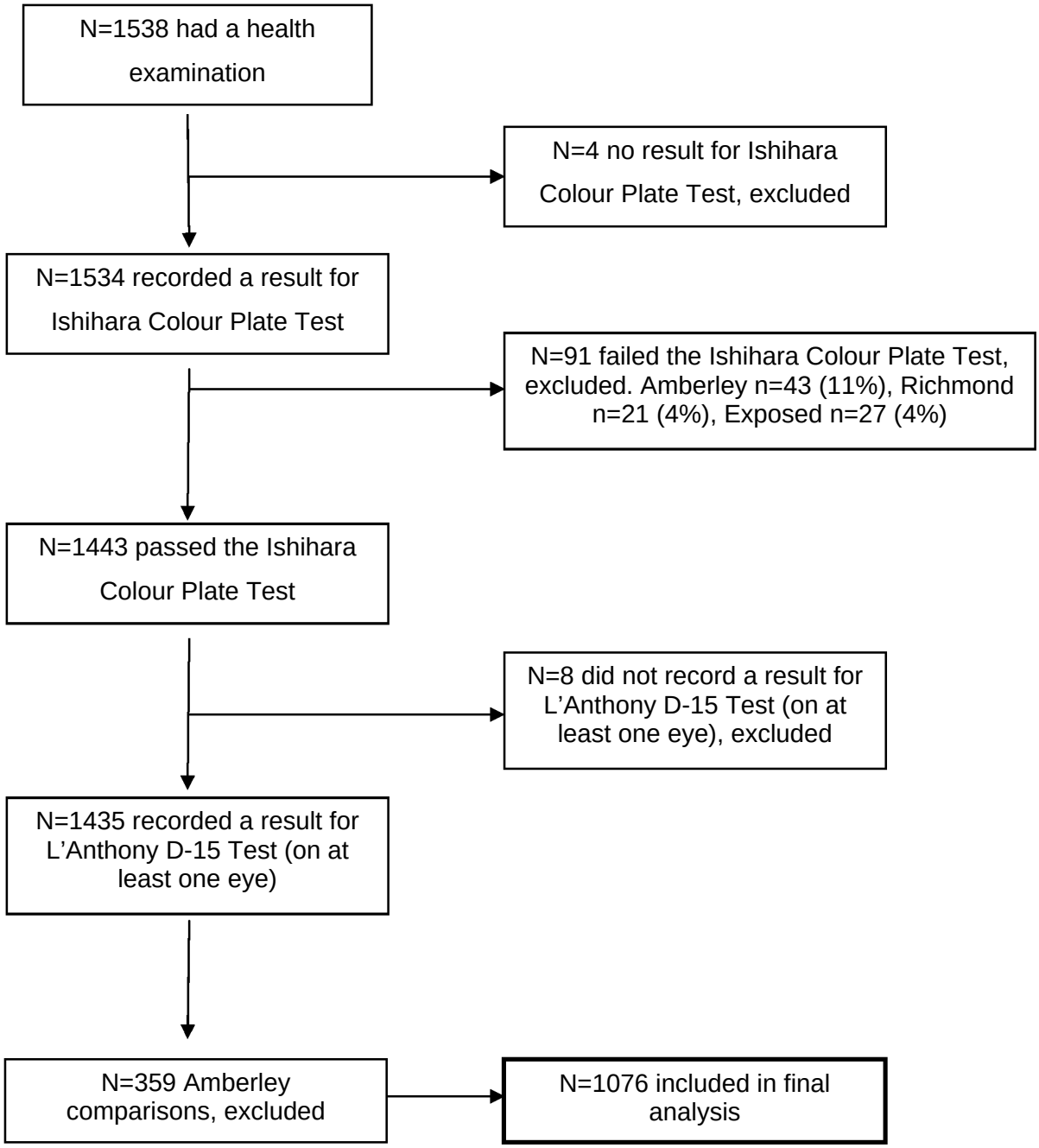
Table 13.13 : Relationship between exposure and self-report of motor symptoms – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.42	0.30	0.59	2	49.82	<.0001
	Richmond vs Exposed	0.34	0.25	0.48	.	.	.
Program 1	Amberley vs Exposed	0.39	0.26	0.58	2	35.00	<.0001
	Richmond vs Exposed	0.33	0.22	0.48	.	.	.
Program 2	Amberley vs Exposed	0.39	0.25	0.61	2	28.54	<.0001
	Richmond vs Exposed	0.32	0.21	0.49	.	.	.
Dose	Mild exposure vs Unexposed	2.41	1.59	3.65	3	49.23	<.0001
	Moderate exposure vs Unexposed	2.39	1.61	3.56	.	.	.
	Prolonged exposure vs Unexposed	3.20	2.19	4.66	.	.	.

13.5.2 Colour vision

Figure 13.4 illustrates participant numbers in the colour vision analysis.

Figure 13.4 : Flow chart of colour vision participants



Of the 1538 SHOAMP participants who received a health examination, 21 were females. Females were retained in the analyses even though no adjustment could be made for gender because of small numbers. Four participants did not take an Ishihara Colour Plate Test and were excluded from the analysis of colour vision. Of the 1534 participants who took the Ishihara test, 91 failed. These individuals and another eight participants who did not take the L-D-15-D were excluded from further analyses.

While only 4% of the Richmond comparison and 4% of the exposed group failed the Ishihara Test, 11% of the Amberley comparison group failed. This was considered to be due to the Airforce Recruitment practices, whereby the Richmond comparison and exposed individuals, because they were technical personnel, should have been screened for red-green colour vision deficiency on enlistment. Because of these baseline differences the Amberley comparison group was excluded from all further colour vision analyses, and only results for the Richmond comparison and exposed groups are reported. There was no heterogeneity across centres (χ^2 = 5.94, df = 7, p = 0.55 for abnormal CCI and χ^2 = 6.12, df = 7, p = 0.53 for clinical diagnosis of abnormal colour vision). Table 13.14 shows the number and percentage of individuals with normal colour vision, slight deficits and moderate deficits for the exposed and Richmond groups.

Table 13.14 : Colour Confusion Index (CCI) severity category by exposure group and dose (unadjusted)

CCI severity category	Richmond N = 516		Exposed N = 616	
	n	%	n	%
Normal	175	*36*	195	*33*
Slight	166	*34*	195	*33*
Moderate/Severe	151	*31*	194	*33*

Overall almost two-thirds of participants in the study had slight, moderate or severe colour vision deficits. Multivariate regression analysis of this dichotomised CCI (normal versus slight/moderate) indicates a borderline decrease in abnormal colour vision in the Richmond group relative to the exposed group (OR 0.74; 95% CI: 0.55, 1.02) (see Table 13.15). This decrease becomes significant in the reduced model (OR 0.73, 95% CI: 0.54, 0.98) and with the robust standard error estimates (OR 0.74, 95% CI: 0.60, 0.92). This decrease is

consistent in the Program 1 and 2 subgroups but not significant (p=0.34, p=0.41 respectively). There was no clear stepwise increase in risk with increasing exposure (OR 0.84, 1.5, 1.5) although the overall dose-effect was significant (p=0.03).

Table 13.15 : Colour Confusion Index (CCI) severity category – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Richmond vs Exposed	0.74	0.55	1.02	1	3.47	0.0624
Program 1	Richmond vs Exposed	0.83	0.56	1.22	1	0.91	0.3393
Program 2	Richmond vs Exposed	0.83	0.54	1.30	1	0.66	0.4182
Dose	Mild exposure vs Unexposed	0.84	0.53	1.34	3	8.87	0.0310
	Moderate exposure vs Unexposed	1.54	1.01	2.35	.	.	.
	Prolonged exposure vs Unexposed	1.54	1.00	2.37	.	.	.

Table 13.16 presents the distribution of clinical diagnoses of colour vision deficits by group. Again, over half of all participants have some form of colour vision deficit.

Table 13.16 : Type of colour vision deficit (clinical diagnosis) by exposure group and dose (unadjusted)

Clinical Diagnosis*	Richmond		Exposed	
	N=516	%	N=616	%
No defect (in both eyes)	223	*46*	253	*44*
Blue-yellow	224	*46*	273	*47*
Red-green	8	*1.7*	19	*3.3*
Mixed	28	*5.8*	37	*6.5*
Non-specific	44	*9.1*	44	*7.7*
Any clinical diagnosis	265	*54*	325	*56*

** Diagnoses are mutually exclusive within but not between eyes; an individual could have one diagnosis in the left eye and a different diagnosis in the right eye.*

The results of the multivariate regression for the dichotomised clinical diagnosis (Table 13.17) indicate there was a borderline decrease in colour vision abnormalities in the Richmond group versus the exposed group (OR 0.79, 95% CI: 0.58, 1.06). This remained borderline in the reduced model (OR 0.77, 95% CI: 0.57, 1.02) but became significant with the robust standard error estimates (OR 0.79, 95% CI: 0.70, 0.88). Results are consistent but not significant in Program 1 and Program 2 subgroups, and there was no clear dose-response effect (p=0.11).

Table 13.17 : Abnormal colour vision (clinical diagnosis) – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Richmond vs Exposed	0.79	0.58	1.06	1	2.52	0.1125
Program 1	Richmond vs Exposed	0.83	0.56	1.23	1	0.87	0.3509
Program 2	Richmond vs Exposed	0.95	0.63	1.42	1	0.07	0.7902
Dose	Mild exposure vs Unexposed	0.89	0.58	1.36	3	5.17	0.1599
	Moderate exposure vs Unexposed	1.35	0.89	2.03			
	Prolonged exposure vs Unexposed	1.39	0.92	2.10			

13.5.3 Olfaction

Of the 1538 individuals who undertook a health examination, one refused to participate in the olfaction examination. 228 individuals were excluded from analyses because they failed the screening criteria, while a further 111 individuals had missing data from some component of the olfaction test. Olfaction score was approximately normally distributed. There was no heterogeneity across HSA centres (F=0.85, p=0.61). The distribution of total olfaction scores across the three exposure groups is shown in Table 13.18. Table 13.19 shows the results of the multiple regression analysis.

Table 13.18 : Univariate analysis for olfaction score by exposure group

Measure	Amberley	Richmond	Exposed
Mean	31.49	31.55	31.53
Standard Deviation	4.54	4.22	4.54
50th Percentile	31.50	31.75	31.75
Lower Quartile	28.75	29.00	28.75
Upper Quartile	34.25	34.50	34.50

Table 13.19 : Relationship between exposure and olfaction score from regression analysis – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	0.10	-0.57	0.77	0.08	2	0.1603
	Richmond	0.58	-0.05	1.21	.	.	.
Program 1	Amberley	0.23	-0.59	1.04	0.09	2	0.1521
	Richmond	0.70	-0.08	1.48	.	.	.
Program 2	Amberley	-0.61	-1.51	0.30	0.07	2	0.3217
	Richmond	-0.18	-1.07	0.70	.	.	.
Dose	Mild exposure	-0.55	-1.40	0.29	0.08	3	0.3783
	Moderate exposure	0.08	-0.74	0.90	.	.	.
	Prolonged exposure	-0.52	-1.34	0.30	.	.	.

There is no statistically significant difference in olfaction scores among the exposure groups in either the primary or secondary analyses and no evidence of a dose-response relationship, although Richmond does appear to score higher than the exposed by ~0.6 point. This remains non-significant in the reduced model but becomes significant with the robust standard error estimates (OR 0.58, 95% CI: 0.16, 1.00). Table 13.20 shows the number and percent of participants reporting increased sensitivity to smells or odours in the past month. The proportion of participants with self-reported increased smell sensitivity appears higher for the exposed group than for both Amberley and Richmond comparison groups. Logistic regression analyses demonstrated statistically significantly lower odds of

self-reported symptoms for Amberley and Richmond comparison groups relative to the exposed group (Table 13.21).

Table 13.20 : Number and percentage of participants with self-reported increased sensitivity to smells in the past month by exposure group

PQ item	Amberley N = 405		Richmond N = 516		Exposed N = 616	
	n	%	n	%	n	%
Self-reported increased sensitivity to smells	36	*7.4*	26	*4.4*	97	*15*

Table 13.21 : Relationship between exposure and self-reported increased sensitivity to smells or odours – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	0.53	0.35	0.82	2	33.56	<.0001
	Richmond	0.26	0.16	0.41	.	.	.
Program 1	Amberley	0.44	0.26	0.74	2	32.03	<.0001
	Richmond	0.22	0.13	0.37	.	.	.
Program 2	Amberley	0.43	0.25	0.75	2	26.62	<.0001
	Richmond	0.22	0.12	0.39	.	.	.
Dose	Mild exposure	2.03	1.17	3.54	3	35.14	<.0001
	Moderate exposure	2.36	1.41	3.96	.	.	.
	Prolonged exposure	3.80	2.40	6.01	.	.	.

The odds ratios for all exposed, Program 1 and Program 2 ranged from 0.43-0.53 for the Amberley comparison group relative to the exposed group, and 0.22 to 0.26 for Richmond compared to the exposed group. There was also evidence of a dose response relationship (OR 2.0, 2.4, 3.8).

13.5.4 Diagnosed or treated medical conditions

Two participants reported having been diagnosed with MND. A further two reported a diagnosis of MS (see Table 13.22). Because the numbers are so small, no further analysis was undertaken.

Table 13.22 : Number and percentage of participants who have been diagnosed with motor neurone disease or multiple sclerosis by exposure group

	Amberley N = 477		Richmond N = 583		Exposed N = 621	
	n	***%***	**n**	***%***	**n**	***%***
Self-report of previous physician diagnosis of motor neurone disease	0	*0*	1	*0.17*	1	*0.16*
	N=478	***%***	**N=582**	***%***	**N=624**	***%***
Self-report of previous physician diagnosis of multiple sclerosis	1	*0.21*	0	*0*	1	*0.16*

13.6 Discussion

13.6.1 Vibration perception threshold (VPT)

Vibration sense was analysed in two ways:

1) as a continuous measure log-transformed to obtain a normal distribution
2) dichotomised according to clinical severity.

Using the transformed continuous measure there was a very slight increase in VPT at the finger in the DSRS exposed group, i.e. poorer sensation. This was significant versus the Richmond comparison group only. Although this change was statistically significant, the magnitude of the difference is small; on a scale of 0 to 50 millivolts (mV), the threshold of the DSRS group was on average 1 mV higher than the comparison groups. The median threshold for the comparison groups was five at the finger, well within the normal cut off of seven. This means that although the threshold was increased by 1 mV in the exposed group, the value was still within the normal range. These results were not seen at the toe. This is puzzling, in that any systemic neurotoxin tends to affect nerve cells with longer axons first, hence one would expect any neuropathy to be more pronounced and severe in the lower limbs rather than the upper limbs. Further weakening this result is the fact that using the clinical categories of normal and abnormal VPT (including mild, moderate, severe and extreme neuropathy) there was no indication of any effect at the finger, and none of the analyses reached statistical significance. There was also no indication of any dose-response relationship.

13.6.2 Self-reported symptoms

The DSRS exposed group had statistically significantly increased odds of self-reported sensory and motor symptoms compared to both comparison groups. Sensory symptoms were increased in the exposed group with an odds ratio of 2.0 versus both the Amberley and Richmond comparison groups (95% CI: 1.5-2.6 and 1.6-2.6, respectively). These estimates were consistent and significant in the Program 1 and 2 subgroups; and there was a weak dose-response effect, with the odds ratio of symptoms increasing from 1.9 to 2.1 from lowest to highest exposed groups.

This pattern was even more marked with the self-reported motor symptoms. The odds ratio was elevated in the exposed group, with a value of 2.4 (95% CI: 1.7-3.3) versus Amberley and 2.9 (95% CI: 2.1-4.0) versus Richmond. These results were consistent and significant in the Program 1 and 2 subgroups; and there was a dose-response effect, with the odds ratio increasing from 2.4 to 2.6 to 3.2 in the lowest to highest exposed groups.

13.6.3 Colour vision

One of the most striking findings was the higher proportion of red-green colour vision deficit in the Amberley comparison group compared to the exposed and Richmond comparisons. Since red-green colour deficits are largely congenital, this reflects the fact that the Air Force used different screening criteria for recruiting technical and non-technical personnel. It is impossible to know to what extent this screening procedure affected blue-yellow colour vision, since entry results on these tests were not available and it is not possible to tell if there were any baseline differences. If differences were detected between the Amberley comparison group and the exposed group it would not be possible to attribute these with any certainty to DSRS activities. Hence the Amberley non-technical group was dropped from this analysis as not being a valid comparator.

There was an increase in impaired colour vision in the exposed group versus the Richmond comparison, with an odds ratio of about 1.3 (i.e. an approximate 30% increase). This increase was more marked when colour vision was assessed using CCI rather than clinical diagnoses, but in both cases the results were only significant in the reduced model or with the robust variance estimates. Program 1 and 2 subgroup estimates were consistent but there was no clear dose-response effect. Over one-half of all participants (both exposed and comparison) had an abnormal Colour Confusion Index or a clinical colour deficit. This proportion seems very high compared to other studies such as those of Gong[60] and Ihrig.[61] However, approximately 60% of SHOAMP participants were aged more than 40 years at the time of colour vision testing, and these results may simply reflect the older age group.

13.6.4 Olfaction

On objective testing, there was no detectable difference in overall smell ability between the exposed and comparison groups; this was consistent in subgroup comparisons of Programs 1 and 2, and in the lack of a dose-response effect. However, there was a strong and statistically significant approximate three-fold increase in self-reported sensitivity to smells,

which was consistent in subgroup analysis and which followed a dose-response curve. Although this latter result may seem paradoxical, odours and irritation are complex subjects. The neurosensory mechanisms of olfaction and common chemical senses provide protection against noxious airborne chemicals. Activation and amplification of these neurosensory mechanisms could explain some of the “sensitivity” experienced by people exposed to airborne irritants and odours.

13.6.5 MS and MND

There were too few cases of Multiple Sclerosis or Motor Neurone Disease to analyse.

13.7 Conclusions

There is a definite approximately two- to three-fold increase in subjective (self-reported) sensory and motor neuropathic symptoms in the DSRS exposed group, relative to both comparison groups. This is consistent, statistically significant, and supported by a dose-response relationship. There is, however, little objective data from the health examination to support this. Most vibration testing showed no difference, and what little difference there was (at the finger and versus Richmond only) translates into a 1mv change in threshold, which is of no clinical significance. There was also a very weak 30% increase in impaired colour vision in the exposed group versus Richmond, but this was only significant in the reduced model or with the robust standard error estimates and there was no dose-response effect. There was no detectable change in olfaction, although there was an approximate three-fold increase in self-reported sensitivity to smells.

13.8 References

1. Knave, B., Persson, H.E., Goldberg, J.M. & Westerholm, P. Long-term exposure to jet fuel: an investigation on occupationally exposed workers with special reference to the nervous system. *Scand J Work Environ Health* **2**, 152-64 (1976).
2. Demers, R.Y., Markell, B.L. & Wabeke, R. Peripheral vibratory sense deficits in solvent-exposed painters. *J Occup Med* **33**, 1051-4 (1991).
3. Malarkey, W.B., Strauss, R.H., Leizman, D.J., Liggett, M. & Demers, L.M. Endocrine effects in female weight lifters who self-administer testosterone and anabolic steroids. *Am J Obstet Gynecol* **165**, 1385-90 (1991).
4. Nasterlack, M. et al. A multidisciplinary cross-sectional study on solvent-related health effects in painters compared with construction workers. *Int Arch Occup Environ Health* **72**, 205-14 (1999).
5. Broadwell, D.K., Darcey, D.J., Hudnell, H.K., Otto, D.A. & Boyes, W.K. Work-site clinical and neurobehavioral assessment of solvent-exposed microelectronics workers. *Am J Ind Med* **27**, 677-98 (1995).
6. Myers, J.E., Nell, V., Colvin, M., Rees, D. & Thompson, M.L. Neuropsychological function in solvent-exposed South African paint makers. *Journal of Occupational and Environmental Medicine* **41**, 1011-1018 (1999).
7. Demers, P.A. et al. Occupational exposure to electromagnetic fields and breast cancer in men. *American Journal of Epidemiology* **134**, 340-347 (1991).
8. Nasterlack, M. et al. A multidisciplinary cross-sectional study on solvent-related health effects in painters compared with construction workers. *International Archives of Occupation and Environmental Health* **72**, 205-14 (1999).
9. Verriest, G. Further study on acquired deficiency of colour discrimination. *Journal of the Optical Society of America* **53**, 185-195 (1963).
10. Mergler, D.B., L. Assessing color vision loss among solvent-exposed workers. *American Journal of Industrial Medicine* **12**, 195-203 (1987).
11. Hart, W.M. Acquired dyschromatopsias. *Survey of Ophthalmology* **32**, 10-31 (1987).
12. Verriest, G. Further study on acquired deficiency of color discrimination. *Journal Optical Society of America* **53**, 185-195 (1963).
13. Mergler, D. & Blain, L. Assessing color vision loss among solvent-exposed workers. *Am J Ind Med* **12**, 195-203 (1987).
14. Gobba, F. Colour vision: a sensitive indicator of exposure to neurotoxins. *Neurotoxicology* **21**, 857-62 (2000).

15. Gobba, F. & Cavalleri, A. Evolution of color vision loss induced by occupational exposure to chemicals. *Neurotoxicology* **21**, 777-81 (2000).
16. Raitta, C., Seppalainen, A.N. & Huuskonen, M.S. N-hexane maculopathy in industrial workers. *Albrecht Von Graefes Arch Klin Exp Ophthalmol* **209**, 99-110 (1978).
17. Chang, Y.C. Neurotoxic effects of n-hexane on the human central nervous system: evoked potential abnormalities in n-hexane polyneuropathy. *J Neurol Neurosurg Psychiatry* **50**, 269-74 (1987).
18. Cavalleri, A., Gobba, F., Nicali, E. & Fiocchi, V. Dose-related colour vision impairment in toluene-exposed workers. *Arch Environ Health* **55**, 399-404 (2000).
19. Muttray, A., Wolters, V., Mayer-Popken, O., Schicketanz, K.H. & Konietzko, J. Effect of subacute occupational exposure to toluene on color vision. *Int J Occup Med Environ Health* **8**, 339-45 (1995).
20. Urban, P. & Lukas, E. Visual evoked potentials in rotogravure printers exposed to toluene. *Br J Ind Med* **47**, 819-23 (1990).
21. Vrca, A., Bozicevic, D., Bozikov, V., Fuchs, R. & Malinar, M. Brain stem evoked potentials and visual evoked potentials in relation to the length of occupational exposure to low levels of toluene. *Acta Med Croatica* **51**, 215-9 (1997).
22. Zavalic, M., Mandic, Z., Turk, R., Bogadi-Sare, A. & Plavec, D. Quantitative assessment of color vision impairment in workers exposed to toluene. *Am J Ind Med* **33**, 297-304 (1998).
23. Zavalic, M. et al. Assessment of colour vision impairment in male workers exposed to toluene generally above occupational exposure limits. *Occup Med (Lond)* **48**, 175-80 (1998).
24. Zavalic, M. et al. Qualitative color vision impairment in toluene-exposed workers. *Int Arch Occup Environ Health* **71**, 194-200 (1998).
25. Zavalic, M., Turk, R., Bogadi-Sare, A. & Skender, L. Colour vision impairment in workers exposed to low concentrations of toluene. *Arh Hig Rada Toksikol* **47**, 167-75 (1996).
26. Blain, L. & Mergler, D. Dyschromatopsia in subjects occupationally exposed to organic solvents. *J Fr Ophtalmol* **9**, 127-33 (1986).
27. Donoghue, A.M., Dryson, E.W. & Wynn-Williams, G. Contrast sensitivity in organic-solvent-induced chronic toxic encephalopathy. *J Occup Environ Med* **37**, 1357-63 (1995).
28. Gonzalez, M., Velten, M. & Cantineau, A. Increased acquired dyschromatopsia among solvent-exposed workers: an epidemiology study on 249 employees of an aluminum-foil printing factory. *Int Arch Occup Environ Health* **71**, 317-24 (1998).

29. Mergler, D., Belanger, S., De Grosbois, S. & Vachon, N. Chromal focus of acquired chromatic discrimination loss and solvent exposure among printshop workers. *Toxicology* **49**, 341-8 (1988).
30. Muttray, A., Wolff, U., Jung, D. & Konietzko, J. Blue-yellow deficiency in workers exposed to low concentrations of organic solvents. *Int Arch Occup Environ Health* **70**, 407-12 (1997).
31. Nolfe, G., Palma, V., Guadagnino, M., Serra, L.L. & Serra, C. Evoked potentials in shoe-workers with minimal polyneuropathy. *Electromyogr Clin Neurophysiol* **31**, 157-62 (1991).
32. Pratt, H., Karim, N., Bleich, N. & Mittelman, N. Short latency visual evoked potentials in occupational exposure to organic solvents. *Neurophysiol Clin* **30**, 306-12 (2000).
33. Semple, S. et al. Impairment of colour vision in workers exposed to organic solvents. *Occup Environ Med* **57**, 582-7 (2000).
34. Druzdz, A., Dabrowicz, K. & Grzybowski, G. Examination of visual evoked potentials in persons exposed to prolonged contact with ethyl benzene vapors. *Klin Oczna* **99**, 47-50 (1997).
35. Mascagni, P., Consonni, D., Bregante, G., Chiappino, G. & Toffoletto, F. Olfactory function in workers exposed to moderate airborne cadmium levels. *Neurotoxicology* **24**, 717-724 (2003).
36. Amoore, J.E. Effects of chemical exposure on olfaction in humans. in *Toxicology of the nasal passages* (ed. Barrows, C.S.) 155-190 (Hemisphere Publishing, Washington, 1986).
37. Gobba, F. Occupational exposure to chemicals and sensory organs: a neglected research field. *Neurotoxicology* **24**, 675-91 (2003).
38. Hotz, P. et al. Smell or taste disturbances, neurological symptoms, and hydrocarbon exposure. *International Archives of Occupational & Environmental Health* **63**, 525-30 (1992).
39. Sandmark, B., Broms, I., Lofgren, L. & Ohlson, C.G. Olfactory function in painters exposed to organic solvents. *Scand J Work Environ Health* **15**, 60-3 (1989).
40. Emmett, E.A. Parosmia and hyposmia induced by solvent exposure. *Br J Ind Med* **33**, 196-8 (1976).
41. Baelum, J., Lundqvist, G.R., Molhave, L. & Andersen, N.T. Human response to varying concentrations of toluene. *International Archives in Occupational and Environmental Health* **62**, 65-71 (1990).
42. Hawkes, C.H. & Fox, A.J. Motor neurone disease in leather workers. *The Lancet* **1**, 507 (1981).

43. Buckley, J. et al. Motor neuron disease in England and Wales, 1959-1979. *Journal of Neurology, Neurosurgery and Psychiatry* **46**, 197-205 (1983).
44. Gunnarsson, L.G., Bodin, L., Soderfeldt, B. & Axelson, O. A case-control study of motor neurone disease: its relation to heritability, and occupational exposures, particularly to solvents. *British Journal of Industrial Medicine* **49**, 791-798 (1992).
45. Candua, S.M., Manzo, L. & Costa, L.G. Role of occupational neurotoxins in psychiatric and neurodegenerative disorders. in *Occupational Neurotoxicology* (ed. Manzo, L.) 131-168 (CRC Press: Boca Raton, 1998).
46. Schulte, P.A., Burnett, C.A., Boeniger, M.F. & Johnson, J. Neurodegenerative diseases: occupational occurrence and potential risk ractors, 1982 through 1991. *American Journal of Public Health* **86**, 1281-1288 (1996).
47. Hensiek, A.E., Sawcer, S.J. & Compston, D.A. Searching for needles in haystacks-the genetics of multiple sclerosis and other common neurological diseases. *Brain Res Bull* **61**, 229-34 (2003).
48. Landtblom, A.M., Flodin, U., Soderfeldt, B., Wolfson, C. & Axelson, O. Organic solvents and multiple sclerosis - a synthesis of the current evidence [Review]. *Epidemiology* **7**, 429-433 (1996).
49. London, L., Nell, V., Thompson, M.L. & Myers, J.E. Effects of long-term organophosphate exposures on neurological symptoms, vibration sense and tremor among South African farm workers. *Scand J Work Environ Health* **24**, 18-29 (1998).
50. Chuang, H.Y. et al. Vibration perception thresholds in workers with long term exposure to lead. *Occup Environ Med* **57**, 588-94 (2000).
51. Manzo, L. & Cosi, V. Evaluation and Management of Neurotoxicity in Occupational Illness. in *Occupational Neurotoxicology* (ed. Manzo, L.) 124 (CRC Press, Boca Raton, 1998).
52. Doty, R.L., Shaman, P., Kimmelman, C.P. & Dann, M.S. University of Pennsylvania Smell Identification Test: a rapid quantitative olfactory function test for the clinic. *Laryngoscope* **94**, 176-8 (1984).
53. Cain, W.S., Gent, J.F., Goodspeed, R.B. & Leonard, G. Evaluation of olfactory dysfunction in the Connecticut Chemosensory Clinical Research Center. *Laryngoscope* **98**, 83-8 (1988).
54. Wolfensberger, M., Schnieper, I. & Welge-Lussen, A. Sniffin'Sticks: a new olfactory test battery. *Acta Otolaryngol* **120**, 303-6 (2000).
55. Pruszewicz, A. Examination of smell and taste. *Otolaryngol Pol* **19**, 29-37 (1965).
56. Hummel, T., Sekinger, B., Wolf, S.R., Pauli, E. & Kobal, G. 'Sniffin' sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chemical Senses* **22**, 39-52 (1997).

57. Apter, A., Mott, A. & Cain, W. Olfactory loss and allergic rhinitis. *Journal of Allergy and Clinical Immunology* **90**, 670-680 (1992).
58. Sullivan, J.B. Olfactory and Nasal Toxicology. (2001).
59. Cornblath, D.R. et al. Total neuropathy score: validation and reliability study. *Neurology* **53**, 1660-4 (1999).
60. Gong, Y. et al. Visual dysfunction in workers exposed to a mixture of organic solvents. *Neurotoxicology* **24**, 703-10 (2003).
61. Ihrig, A., Nasterlack, M., Dietz, M.C., Hoffmann, J. & Triebig, G. Pilot study on prevalence of color vision dysfunction in long-term solvent-exposed painters. *Ind Health* **41**, 39-42 (2003).

This page has been left blank intentionally
for formatting purposes

14 Sexual Function and Reproductive Health

Chapter summary

This chapter details the measures and results for male sexual function (specifically erectile function) and female reproductive health (including pregnancy results and difficulties getting pregnant). The 15-item International Index of Erectile Function for males identified significantly lower levels of self-reported sexual functioning in the DSRS exposed group, with a two-fold increase in erectile dysfunction in the exposed versus the comparison groups for the overall and subgroup analyses, and with a significant dose-response effect. The exposed group were also significantly more likely to self-report a “loss of interest in sex” and “problems with sexual function” than were the comparison groups. Female DSRS workers and female partners of male study participants were asked to complete a Female Reproductive Questionnaire. There was no statistically significant evidence in the current study of any association between having a male partner with DSRS exposure and miscarriage or stillbirth; this finding was consistent in subgroup analyses and lacked a dose-response effect. There were also no detectable differences in reported difficulties getting pregnant or in seeing a fertility specialist.

Chapter contents

14.1 Introduction

The F-111 DSRS Interim Health Care Scheme had received complaints from DSRS personnel regarding adverse effects on sexual functioning. Concern had also been raised about the possibility of birth defects in the children of workers in the F-111 DSRS program – both for females who were involved in DSRS activities as part of their Air Force employment and for female partners of male DSRS workers who may have been exposed to chemical substances brought back on their partner's work clothing. In response to these concerns, measures of male and female sexual/reproductive function were included as part of the General Health and Medical Study.

Two recent reviews of occupational exposures on the reproductive system[1,2] highlight organic solvents, ethylene glycol esters and aromatic hydrocarbons as major toxins, all of which are relevant to F-111 DSRS. These effects are summarised in Table 14.1.

Table 14.1 : Reproductive dysfunction due to solvent exposure

Names of the solvent	Observed effects - Males	References
2-ethoxy ethanol	Lower total sperm count	Welch et al.[3,4]
Exposure to aromatic solvent	Reduced semen quality	Tielemans et al.[5]
Names of the solvent	**Observed effects - Females**	**References**
Exposure to organic solvent in petrochemical industry	Maternal exposure leads to reduced birth weight	Ha et al.[6]
Exposure to ethylene glycol ether	Female sub-fertility	Chen et al.[7]
Toluene	Reduced fecundity	Pleng-Bonig and Karmaus[8]

Adapted from: Kumar, S. Occupational exposure associated with reproductive dysfunction. Journal of Occupational Health 46, 1-19 (2004).

Most studies in males report adverse affects on sperm density, total sperm count and motility[5,9-11] but not on erectile function. Erectile dysfunction was assessed by Oliva et al.,[12] who reported that exposure to solvents significantly increased the risk of having a "flat

erectile pattern" (i.e. no erectile activity during their study of nocturnal erections), with an odds ratio of 12.2 (p=0.03, CI: 1.2, 124.8).

Evidence has also accumulated that hydrocarbons in fuels and solvents are reproductive toxicants: reductions in female fertility have been identified in occupational groups exposed to organic solvents containing benzene,[13] toluene,[8] and mixtures of solvents.[14,15] Of relevance to SHOAMP are studies of mixed solvent exposure, which report male-mediated effects such as increased spontaneous abortions and congenital malformations where the fathers were exposed to fuels, toluene and xylene.[16,17] An increased risk of low birth weight for infants was also reported by Daniell and Vaughan[18] with regard to paternal employment in solvent-exposed occupations such as body shop workers (RR 1.6, CI: 1.1-2.4) or painters (RR 1.4, CI: 0.9-2.1). Similarly, McDonald et al.[16] found an increase in the risk of foetal loss when the male partner was employed as a mechanic or repairer. Yet some associations reported between maternal exposure to organic solvents and reduced birth weight[6] and reduced fecundity[8], also find no association with paternal exposure in the same circumstances.

14.2 Measures

Regarding the assessment of male sexual function, the protocol for the General Health and Medical Study did not include provision for the collection of sperm samples. Although physiologically-based diagnostic procedures may be used, it was considered that participation rates could have been adversely affected if such methods had been included as part of the health examination, particularly as sexual function can be appropriately assessed by self-report techniques.[19] In order to assess self-perceived erectile function in SHOAMP participants, the 15-item International Index of Erectile Function[20] scale (IIEF) was included in the male version of the Postal Questionnaire for self-completion by consenting participants from both the exposed and comparison groups. Data were self-reported, and returned to the study team by post (for those individuals who did not consent to a health examination) or delivered to HSA personnel at the time of the health examination. Although other instruments exist for male sexual function, several limitations have been identified regarding their use in the identification of erectile function, such as excessive number of items for the respondent to complete, narrow focus, and lack of validation.[21-23]

Two items from the Postal Questionnaire (but separate to the IIEF) were also included for analysis: item 2.42, self-reported “loss of interest in sex”; and item 2.43, “problems with sexual functioning”, both with a yes/no response.

Two additional areas of investigation originally included in the terms of reference regarding the incidence of major congenital abnormality and malignancies in children were not pursued as part of the Health Study. Due to concerns about children being subjected to the same level of detailed medical examinations as other participants, the issue of congenital abnormalities was not included as part of the health examination. Instead, the issue of maternal health, including miscarriage and pregnancy difficulties, was explored via a specially-developed Reproductive Questionnaire. It has been established that male partners are not as reliable a source of information as females for these types of outcomes; men tend to mis-report the timing of events and to under-report low birth weight, spontaneous abortions and induced abortions.[24] Therefore, this questionnaire was directed to female DSRS workers and to female partners of male study participants.

The questionnaire asked about the result of any pregnancies experienced during the period of F-111 DSRS involvement, age at time of conception, and health habits during each pregnancy. Where a pregnancy was recorded, each person was then asked if they had encountered problems getting pregnant and if a specialist was sought for fertility problems. Male participants in the exposed and comparison cohorts were sent the Reproductive Questionnaire as part of their study invitation package. The questionnaire was sealed in a specially-addressed envelope that instructed the male participant to forward the envelope to a current or past female partner for her to complete, and explained that more than one female could receive a questionnaire. It was up to the male as to whether the envelope was passed on, and then up to the female recipient of the questionnaire as to whether or not she completed and returned it. The study team had no knowledge of how many female partners actually received the Reproductive Questionnaire.

14.3 Potential confounders

For male sexual functioning, age category, posting category and rank category were included as potential confounders in multiple analyses. Additional potential confounders were civilian exposures to organic solvents, lead, smoking, BMI, and depression and anxiety (as measured by the CIDI; see Chapter 8). Females were excluded from these analyses.

Potential confounders of interest for female reproductive outcomes included the age of the female respondent at each pregnancy, the male partner's rank and posting category, whether the female respondent indicated that alcohol was consumed at all during the pregnancy, whether she had smoked at all during the pregnancy, and whether any type of beverage containing caffeine had been consumed during the pregnancy. Only female participants or partners were eligible for these analyses.

14.4 Analyses

All analyses generally followed the outline described in Chapter 6 (Analysis). However, item analysis, including exploratory factor analysis, was undertaken for the IIEF prior to the usual approach.

14.4.1 Male sexual function

The IIEF required detailed responses to 15 items on 5 and 6 point Likert-type scales. The 15 items were the complete set from the International Index of Erectile Function originally reported by Rosen et al.[20] Respondents were asked to complete the items with reference to the previous four weeks. Items 1–10 were scaled on a six-point scale, and items 11–15 were scaled on a 5-point scale except item 14 where an additional choice "no current partner" was included. Rosen et al.[20] identified 5 domains on the IIEF, constructed from the 15 items. The domains identified were erectile function (6 items), orgasmic function (2 items), sexual desire (2 items), intercourse satisfaction (3 items), and overall satisfaction (2 items). Scales for each domain are normally produced by adding together scores for items relating to each domain. In addition, items 1, 2, 3, 4, 5 and 15 represent the domain of "erectile function"; a cut-off score of 25 or less, out of 30, provides an indication of clinically significant erectile

dysfunction.[25] The scales were constructed, correlated and compared to those of Rosen et al.[20] Item analysis was undertaken across the whole 15 items, and factor analysis was undertaken to compare with the factors defined by Rosen et al.

The entire 15 items from the IIEF were combined to form one scale, and mean scores were compared by multiple analysis of variance. The erectile function scale was dichotomised and proportions were compared by multiple logistic regression.

14.4.2 Female reproductive health

Analysis of female reproductive health was undertaken using the STATA statistical package.[26] Pregnancies reported by female partners of male participants and by female DSRS workers were referenced to a posting date. The five posting categories were: 1975-79, 1980-84, 1985-89, 1990-94 and 1995-99. Events recorded as occurring prior to the posting period of interest were excluded from analyses. Events recorded as occurring after the final cut-off date* were excluded from analyses. Pregnancy results were dichotomised as “live birth” versus “other” (which included stillbirth or miscarriage). The chi-square statistic was used to compare the proportion of females who had reported difficulty getting pregnant, and the proportion who reported seeing a specialist, across the three groups. Analyses were performed using multiple logistic regression, adjusting for potential confounders and multiple events per participant (i.e. more than one pregnancy result) using the CLUSTER option in STATA. Paternal age was described but not included as a potential confounder due to collinearity with maternal age.

* The date 7/10/2000 which was forty weeks after the suspension of the Spray Seal DSRS Program.

14.5 Results

14.5.1 Male sexual function

General items: loss of interest in sex

Only 12 participants had missing information for this variable, leaving 1680 for analysis. Respondents to the Postal Questionnaire were asked whether in the last month they had experienced a loss of interest in sex. A “yes” response was given by 38% of the exposed group and by 22% of both the Richmond and Amberley comparison groups (see Table 14.2).

Table 14.2 : General description of Postal Questionnaire item “loss of interest in sex”

PQ item	Amberley		Richmond		Exposed	
	N=482	%	N=582	%	N=616	%
Reported loss of interest in sex	105	*22*	126	*22*	234	*38*

Multiple logistic regression analysis demonstrated a strongly significant difference between these groups, p<0.0001 (see Table 14.3). Specifically, there were markedly lower odds of self-reported loss of interest in sex at 0.50 for Amberley (95% CI: 035, 0.70) and 0.57 for Richmond (95% CI: 0.42, 0.78) when compared with the exposed group. Virtually identical effects were observed for Program 1 and 2; these were also significant, as was the reduced model. All dose levels demonstrated elevated odds compared with those who were unexposed; however, there was no clear trend of increasing odds with increasing dose. Specifically, the odds for the lowest to highest exposure groups were 1.6 (95%CI: 1.0, 2.5), 2.1 (95%CI: 1.4, 3.2), and 1.9 (95%CI: 1.3, 2.8).

Table 14.3 : PQ item "loss of interest in sex" – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.50	0.35	0.70	2	20.01	<.0001
	Richmond vs Exposed	0.57	0.42	0.78	.	.	.
Program 1	Amberley vs Exposed	0.55	0.36	0.82	2	9.22	0.0099
	Richmond vs Exposed	0.62	0.43	0.91	.	.	.
Program 2	Amberley vs Exposed	0.51	0.33	0.81	2	8.61	0.0135
	Richmond vs Exposed	0.58	0.38	0.90	.	.	.
Dose	Mild exposure vs Unexposed	1.60	1.04	2.45	3	20.48	0.0001
	Moderate exposure vs Unexposed	2.14	1.44	3.18	.	.	.
	Prolonged exposure vs Unexposed	1.89	1.27	2.82	.	.	.

General items: problems with sexual functioning

Fourteen participants had missing information, leaving 1678 for analysis. Respondents to the Postal Questionnaire were asked to indicate whether they had experienced problems with sexual functioning in the past month. "Yes" responses were given by 32% of the exposed group, 19% of the Amberley group, and 16% of the Richmond group (see Table 14.4).

Table 14.4 : General descriptions of PQ item "problems with sexual function"

PQ item	*Amberley*		*Richmond*		*Exposed*	
	N=483	***%***	**N=582**	***%***	**N=613**	***%***
Reported problems with sexual functioning	93	*19*	91	*16*	197	*32*

Multiple logistic regression analysis of self-reported problems with sexual functioning comparing all exposed with the two comparison groups, demonstrated a strongly significant difference between these groups, p<0.0001 (see Table 14.5).

Table 14.5 : PQ item "problems with sexual function" – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.52	0.36	0.76	2	24.95	<.0001
	Richmond vs Exposed	0.43	0.30	0.61	.	.	.
Program 1	Amberley vs Exposed	0.49	0.32	0.75	2	20.79	<.0001
	Richmond vs Exposed	0.39	0.26	0.59	.	.	.
Program 2	Amberley vs Exposed	0.64	0.39	1.06	2	7.23	0.0269
	Richmond vs Exposed	0.51	0.31	0.83	.	.	.
Dose	Mild exposure vs Unexposed	1.75	1.09	2.80	3	21.84	<.0001
	Moderate exposure vs Unexposed	2.50	1.62	3.87	.	.	.
	Prolonged exposure vs Unexposed	1.88	1.21	2.91	.	.	.

Specifically, there were markedly lower odds of self-reported problems with sexual functioning of 0.52 for Amberley (95% CI: 036, 0.76) and of 0.43 for Richmond (95% CI: 0.30, 0.61) when compared with the exposed. Again results were similar and significant for both Program 1 and 2 subgroups (p<0.0001 and p=0.03 respectively) and for the reduced model. All dose levels demonstrated elevated odds compared with those who were unexposed; however, there was no clear trend of increasing odds with increasing dose. Specifically, the odds increased from 1.75 (95% CI 1.09, 2.8) to 2.5 (95% CI 1.6-3.9) to 1.9 (95% CI 1.2, 2.9) with increasing exposure.

International Index of Erectile Function (IIEF)

There were 106 participants who were missing all or some answers for the IIEF. A further 119 participants indicated they did not currently have a partner. Since many of the questions on the IIEF referred to sexual intercourse with a partner, these 119 were removed, leaving 1477 observations for analysis. Included among these valid observations were 95 respondents who reported no sexual activity during the four weeks prior to completing the questionnaire. These respondents were provided with a score of 0 for each question where they indicated no sexual activity. Scores for each of the five domains – erectile function (6 items), orgasmic function (2 items), sexual desire (2 items), intercourse satisfaction (3 items), and overall satisfaction (2 items) – were constructed by adding together the appropriate items. They were then correlated to compare with correlations from Rosen et al. The comparison is presented in Table 14.6, where it can be seen that correlations ranged between 0.56 and 0.86 and were higher than Rosen's original correlations.

Table 14.6 : Pearson correlation coefficients for domains of the IIEF

Sexual Function Domains	EF	OF	SD	IS	OS
EF - Erectile function	1.00000				
Rosen EF values	*1.00*				
OF - Orgasmic function	0.82294	1.00000			
Rosen OF values	*0.55*	*1.00*			
SD - Sexual desire	0.63531	0.62034	1.00000		
Rosen SD values	*0.30*	*0.39*	*1.00*		
IS - Intercourse satisfaction	0.86296	0.73993	0.66390	1.00000	
Rosen IS values	*0.76*	*0.47*	*0.35*	*1.00*	
OS - Overall satisfaction	0.65946	0.56368	0.60769	0.69986	1.00000
Rosen OS values	*0.60*	*0.53*	*0.37*	*0.53*	*1.00*

Adapted from: Rosen et al. The International Index of Erectile Function (IIEF): A multi-dimensional scale for assessment of erectile dysfunction. Urology 1997;49:822-830.

Cronbach's alpha, a measure of internal consistency, was very high when all 15 items were considered as a group, being 0.97 overall and ranging between 0.96 and 0.97 over the set of item deletions. All item versus total correlations were also very high, ranging from 0.66 to 0.91 (see Appendix 14U). This indicates that the 15 questions are measuring a similar entity and is contrary to the notion of separate domains. This was explored further by factor analysis. The results of a factor analysis indicate that, unlike the clinical populations of Rosen et al.,[20] in our population there is no clear factor structure and no clearly discernible domains.

Consequent to the item and factor analysis, it was considered reasonable to analyse the data in two ways:

- all items summed together would be analysed as a continuous variable to provide an indication of overall self-reported sexual function
- the ED domain would be extracted and summed together, dichotomised according to clinical relevance[25] and analysed as a dichotomous variable.

Overall IIEF sexual function

The sum of all 15 items on the IIEF produced a score of self-reported sexual function. This resulted in a distribution highly skewed to the left with values ranging between 5 and 75. Preliminary analysis of variance revealed that the residual errors were not normally distributed and thus the data required transformation. In this case a two-step numerical transformation was employed. Firstly, the direction of the skew was reversed from left to right by taking the maximum value, adding one and then subtracting the actual value. Secondly, the natural logarithm of the value was obtained (see equation below):

$$x = \log((\max+1) - y)$$

where *x* is the transformed value and *y* is the value to be transformed. The effect of this transformation on the mean and median can be observed in Table 14.7. Prior to transformation, each mean is substantially lower than the median, but after transformation the means and medians are closely aligned. The effect of the reversal can be seen where prior to transformation the mean of the exposed is considerably lower than the means for the Amberley and Richmond comparison groups, and after transformation the mean of the exposed is higher, reflecting the reversal of the distribution. Additionally, the standard deviation is reduced proportionally.

Table 14.7 : Summary of overall self-reported sexual function with means and transformed means

	Amberley N=415			Richmond N=530			Exposed N=532		
	Mean	**SD**	**Median**	**Mean**	**SD**	**Median**	**Mean**	**SD**	**Median**
All questions combined	61.39	16.75	68	60.65	16.91	67	54.77	20.08	63
All questions combined and transformed	2.11	1.11	2.08	2.21	1.06	2.20	2.55	1.09	2.56

Multiple linear regression analyses indicated a significant group effect for the overall analysis (p<0.0001) (see Table 14.8) as well as for Program 1 and Program 2 (p<0.0001 and p=0.01 respectively). Amberley scored 0.33 log units lower (i.e. better function) than the exposed (95% CI -0.48, -0.19), and Richmond scored 0.26 log units lower (95% CI –0.39, -0.13). This was similar in the subgroup analyses and in the reduced model. Although the dose-effect was significant, there was no stepwise increase in risk with increasing exposure, with log unit scores increasing from 0.14 to 0.38 to 0.28.

Table 14.8 : Overall sexual function – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Estimate	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	-0.33	-0.48	-0.19	0.21	2	<.0001
	Richmond vs Exposed	-0.26	-0.39	-0.13	.	.	.
Program 1	Amberley vs Exposed	-0.39	-0.58	-0.21	0.23	2	<.0001
	Richmond vs Exposed	-0.31	-0.48	-0.14	.	.	.
Program 2	Amberley vs Exposed	-0.30	-0.49	-0.10	0.19	2	0.0111
	Richmond vs Exposed	-0.23	-0.41	-0.04	.	.	.
Dose	Mild exposure vs Unexposed	0.14	-0.05	0.33	0.20	3	<.0001
	Moderate exposure vs Unexposed	0.38	0.20	0.56	.	.	.
	Prolonged exposure vs Unexposed	0.28	0.11	0.46	.	.	.

IIEF erectile dysfunction

The erectile dysfunction scale was calculated by summing the values of items 1, 2, 3, 4, 5 and 15 from the IIEF, which produced a total score ranging between 1 and 30. This score was dichotomised according to the method of Cappelleri et al.[25], where those scoring ≤25

were classified as having erectile dysfunction, while those scoring >25 were classified as not having erectile dysfunction. The distribution of erectile dysfunction by group is reported in Table 14.9, where 24% of the Richmond and Amberley comparison groups were classified as having self-reported erectile dysfunction, whereas 39% of the exposed group were so classified.

Table 14.9 : Description of erectile dysfunction from IIEF

IIEF within Postal Questionnaire	***Amberley***		***Richmond***		***Exposed***	
	N=415	***%***	**N=530**	***%***	**N=532**	***%***
Erectile dysfunction	99	*24*	127	*24*	206	*39*

Multiple logistic regressions for all exposed and Program 1 and 2 subgroups were all significant (see Table 14.10). Specifically, in the primary analysis, odds ratios of 0.54 were observed for both Amberley (95% CI: 0.37, 0.77) and Richmond (95% CI: 0.38, 0.75). Similar, and statistically significant, results were seen for Program 1 and Program 2 (p=0.0006 and p=0.005 respectively). There appeared to be a trend of increasing odds ratios associated with increasing dose where the odds of self-reported erectile dysfunction associated with mild exposure were 1.2 (95% CI: 0.7, 1.9), the odds associated with moderate exposure were 1.9 (95% CI: 1.2, 2.9), and the odds associated with prolonged exposure were 2.3 (95% CI: 1.5, 3.5).

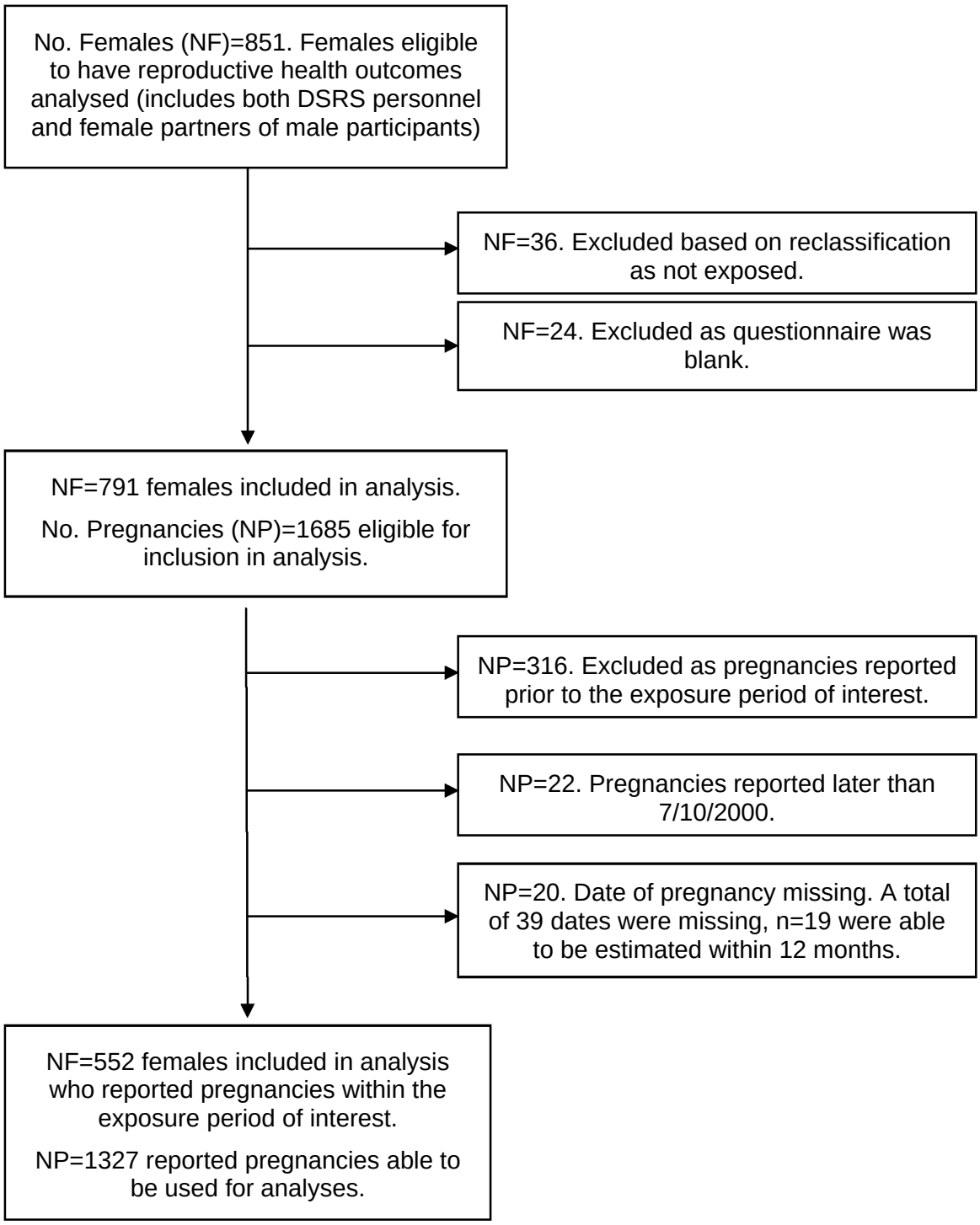
Table 14.10 : IIEF erectile dysfunction: summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.54	0.37	0.77	2	17.32	0.0002
	Richmond vs Exposed	0.54	0.38	0.75	.	.	.
Program 1	Amberley vs Exposed	0.47	0.30	0.74	2	14.78	0.0006
	Richmond vs Exposed	0.48	0.32	0.72	.	.	.
Program 2	Amberley vs Exposed	0.49	0.30	0.79	2	10.77	0.0046
	Richmond vs Exposed	0.48	0.30	0.76	.	.	.
Dose	Mild exposure vs Unexposed	1.17	0.72	1.89	3	19.42	0.0002
	Moderate exposure vs Unexposed	1.91	1.23	2.95	.	.	.
	Prolonged exposure vs Unexposed	2.29	1.50	3.51	.	.	.

14.5.2 Female reproductive health

There were 791 female participants whose data were potentially available for analysis: 24 female DSRS personnel, and 767 female partners of male DSRS personnel (see Figure 14.1).

Figure 14.1 : Flow chart of female reproductive questionnaire respondents and pregnancies



Thirty-six females were excluded due to the reclassification of male partners from their original DVA category 1 “exposed” status to “not exposed”, based on information recorded on their Exposure Questionnaire. Twenty-four females were excluded because they had returned, but not answered, any part of the Female Reproductive Questionnaire. As the current study did not ask male participants to confirm if they had passed on the Female Reproductive Questionnaire, no final numbers are available regarding the possible number of female partners who could have received and completed the questionnaire compared with those who received the questionnaire but did not respond. A total of 1685 pregnancies were reported. Those pregnancies excluded from analyses were: 316 that occurred prior to the exposure period of interest; 22 that occurred later than 7 October 2000 (which was just over nine months from the end of DSRS activities for the purpose of the study); and 20 where the date of the pregnancy was not noted and was not able to be estimated. Consequently, there were 1327 pregnancies available for analysis from 552 female respondents (Figure 14.1).

For pregnancies overall, there were 1072 live births (80%), 20 stillbirths (1.5%), and 235 miscarriages (18%). Unadjusted proportions did not seem to differ greatly between groups, with stillbirths or miscarriages occurring in 17% of births for Amberley, 20% of births for Richmond, and 20% of births for the exposed (Table 14.12). Analysis by multiple logistic regression showed no association with group for all exposed (p=0.54), for Program 1 (p=0.50) or for Program 2 (p=0.34). There was also no dose-response effect (p=0.99) (Table 14.13).

Table 14.11 : Description of pregnancies by exposure group

	Amberley		Richmond		Exposed	
	N	***%***	**N**	***%***	**N**	***%***
Total Respondents	**204**		**294**		**293**	
Women who reported pregnancies within the exposure period of interest	143	*70*	203	*69*	206	*70*
	N	**Mean per female *Median***	**N**	**Mean per female *Median***	**N**	**Mean per female *Median***
Total number of pregnancies reported within the exposure period of interest	351	*2.5* *2*	492	*2.4* *2*	484	*2.3* *2*
	N	**Mean per female *Min-max***	**N**	**Mean per female *Min-max***	**N**	**Mean per female *Min-max***
Maternal age * 21 unknown		*27.2* *(17-44)*		*28.0* *(18-41)*		*27.4* *(16-46)*
Paternal age * 22 unknown		*29.3* *(17-56)*		*30.3* *(21-50)*		*29.2* *(14-48)*

** Data includes 22 pregnancies for 11 female service personnel (24 female service personnel respondents)*

Table 14.12 : Results of pregnancy and exposure group

Totals	**Amberley N=351**		**Richmond N=492**		**Exposed N=484**	
	n	***%***	**n**	***%***	**n**	***%***
Live birth	293	*84*	394	*80*	385	*80*
Stillbirth	6	*1.7*	9	*1.8*	5	*1.0*
Miscarriage	52	*15*	89	*18*	94	*19*
Stillbirth or Miscarriage	58	*17*	98	*20*	99	*20*

Table 14.13 : Pregancies – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	1.13	0.75	1.72	2	1.23	0.54
	Richmond vs Exposed	0.92	0.65	1.30			
Program 1	Amberley vs Exposed	1.24	0.79	1.96	2	1.40	0.50
	Richmond vs Exposed	1.01	0.68	1.51			
Program 2	Amberley vs Exposed	0.87	0.50	1.51	2	2.18	0.34
	Richmond vs Exposed	0.71	0.43	1.17			
Dose	Mild exposure vs Unexposed	0.95	0.56	1.62	3	0.12	0.99
	Moderate exposure vs Unexposed	1.05	0.68	1.62			
	Prolonged exposure vs Unexposed	1.03	0.62	1.72			

In addition to questions about pregnancy outcomes, respondents were asked to record any difficulties getting pregnant and whether or not they had visited a specialist for fertility problems. Formal analysis was not possible as information about key potential confounders such as maternal age at the time of problems or specialist visit was not collected. Overall, of the women responding who reported a pregnancy, 30% of women in the exposed group reported difficulties getting pregnant compared with 27% from Richmond and 21% from Amberley, while 14% of women in the exposed group reported seeing a specialist compared with 18% of the Richmond group and 11% of women in the Amberley group (Table 14.14). The proportions were not significantly different using chi-squared analysis (p=0.18 for difficulties getting pregnant, and p=0.21 for seeing a specialist).

Table 14.14 : Additional questions regarding fertility problems

Reproductive Questionnaire items	Group						Chi-square	df	p-value
	Amberley		Richmond		Exposed				
	N=159	%	N=255	%	N=243	%			
Reported "Difficulties in getting pregnant"	34	*21*	69	*27*	72	*30*	3.38	2	0.18
Reported "ever seen a Specialist about fertility problems"	18	*11*	45	*18*	35	*14*	3.17	2	0.21

14.6 Discussion

The impetus for including measures of male sexual function in this study was the fact that the F-111 DSRS Interim Health Care Scheme had received numerous submissions from male DSRS participants in relation to loss of sexual function. Our current findings are congruent with these spontaneous complaints, i.e. that DSRS activities are associated with reported loss of sexual function in males. All analyses demonstrated a highly statistically significant decrease in sexual function in the exposed group compared to the two comparison groups.

The exposed group had significantly more self-reported loss of interest in sex, with an odds ratio of 2.4 for the exposed group when compared to the Amberley (95%CI: 1.4, 2.9), and 1.8 for the exposed when compared to the Richmond (95%CI: 1.3, 2.4). Similarly, the exposed group had significantly higher odds of self-reported problems with sexual function, the odds ratio being 1.9 for the exposed group when compared to the Amberley (95% CI: 1.3, 2.8), and 2.3 for the exposed when compared to the Richmond (95% CI: 1.6, 3.3). These associations remained significant and were consistent with the secondary analyses for Programs 1 and 2, and both showed a weak dose-response effect. This self-reported decrease in sexual interest and function was corroborated by the results of the IIEF, a validated questionnaire. Results were similar for the overall IIEF score analysed continuously or just for the erectile function domain analysed dichotomously according to a clinical cut-off score.

Overall scores were lower (i.e. better function) for Amberley and Richmond comparison groups, and this was mirrored in the Program 1 and 2 subgroup analyses. Obtaining a poor score on the erectile function domain was 1.9-fold more common in the exposed group compared to both Amberley (95% CI: 1.3, 2.7) and Richmond (95% CI: 1.3, 2.6), and this was also consistent in the Program 1 and 2 subgroups. The latter also showed a clear dose-response effect.

There was no evidence of any association between having a partner with DSRS exposure and miscarriage or stillbirth. This lack of any effect was consistent in subgroup analyses of Programs 1 and 2, and there was no detectable dose-response effect. There was also no detectable difference in the proportions of participants reporting difficulties in getting pregnant, or seeing a fertility specialist. While these results are reassuring, they do not directly address the concerns expressed regarding birth defects or childhood health in the offspring of participants.

It was not possible within the scope of this study to canvas all possible outcomes in this area (e.g. collecting sperm samples from participants, or examining/testing children), or to capture other adverse pregnancy outcomes such as pre-eclampsia. Instead, focus was on those outcomes which could be assessed reliably, and which were common, and which would have the greatest statistical power to detect an effect.

14.7 Conclusions

Using both *ad hoc* and validated questions, the current study showed that there was an average two-fold increased risk of sexual dysfunction, and particularly erectile dysfunction, in exposed males compared to either the Amberley or Richmond comparison groups. This relationship is seen in Programs 1 and 2, and with respect to erectile function it follows a dose-response curve. In terms of female reproductive health, there was no detectable effect of DSRS on miscarriages and stillbirths.

14.8 References

1. Kumar, S. Occupational exposure associated with reproductive dysfunction. *Journal of Occupational Health* **46**, 1-19 (2004).
2. Tas, S., Lauwerys, R. & Lison, D. Occupational hazards for the male reproductive system. *Critical Reviews* **26**, 261-307 (1996).
3. Welch, L.S., Schrader, S.M., Turner, T.W. & Cullen, M.R. Effects of exposure to ethylene glycol ethers on shipyard painters: II. Male reproduction. *American Journal of Industrial Medicine* **14**, 509-26 (1988).
4. Welch, L.S., Plotkin, E. & Schrader, S. Indirect fertility analysis in painters exposed to ethylene glycol ethers: sensitivity and specificity. *American Journal of Industrial Medicine* **20**, 229-40 (1991).
5. Tielemans, E. et al. Occupationally related exposures and reduced semen quality: A case control study. *Fertility and Sterility* **71**, 690-696 (1999).
6. Ha, E. et al. Parental exposure to organic solvents and reduced birth weight. *Archives of Environmental Health* **57**, 207-214 (2002).
7. Chen, P.C., Hsieh, G.Y., Wang, J.D. & Cheng, T.J. Prolonged time to pregnancy in female workers exposed to ethylene glycol ethers in semiconductor manufacturing. *Epidemiology* **13**, 191-196 (2002).
8. Plenge-Bonig, A. & Karmaus, W. Exposure to toluene in the printing industry is associated with subfecundity in women but not in men. *Occupational and Environmental Medicine* **56**, 443-448 (1999).
9. Xiao, G., Pan, C., Cai, Y., Lin, H. & Fu, Z. Effect of benzene, toluene, xylene on the semen quality and the function of accessory gonad of exposed workers. *Industrial Health* **39**, 206-210 (2001).
10. Takahashi, K., Hanoka, T. & Pan, G. Male reproductive health in relation to occupational exposure to endocrine disrupting and other potent chemicals, a review of the epidemiologic literature. *J. UOEH (University of Occupational and Environmental Health)* **26**, 23-40 (2004).
11. Xiao, G., Pan, C., CAI, Y., Lin, H. & Fu, Z. Effect of benzene, toluene, xylene on the semen quality of exposed workers. *Chinese Medical Journal* **112**, 709-712 (1999).
12. Oliva, A., Giami, A. & Multigner, L. Environmental agents and erectile dysfunction: a study in a consulting population. *Journal of Andrology* **23**, 546-550 (2002).
13. Cho, S.I. et al. Benzene exposure and time to conception. *American Journal of Epidemiology* **151**, S13 (2000).

14. Sallmen, M. et al. Reduced fertility among women exposed to organic solvents. *American Journal of Industrial Medicine* **27**, 699-713 (1995).
15. Smith, E.M., Hammonds-Ehlers, M., Clark, M.K., Kirchner, H.L. & Fuortes, L. Occupational exposures and risk of female infertility. *Journal of Occupational and Environmental Medicine* **39**, 138-147 (1997).
16. McDonald, A.D. et al. Fathers' occupation and pregnancy outcome. *British Journal of Industrial Medicine* **46**, 329-333 (1989).
17. Taskinen, H., Anttila, A., Lindbohm, M., Sallmen, M. & Hemminki, K. Spontaneous abortions and congenital malformations among the wives of men occupationally exposed to organic solvents. *Scandinavian Journal of Work and Environmental Health* **15**, 345-352 (1989).
18. Daniell, W.E. & Vaughan, T.L. Paternal employment in solvent related occupations and adverse pregnancy outcomes. *British Journal of Industrial Medicine* **45**, 193-197 (1988).
19. Anderson, B.L. & Broffit, B. Is there a reliable and valid self-report measure of sexual function? *Archives of Sexual Behaviour* **17**, 509-525 (1988).
20. Rosen, R.C. et al. The International Index of Erectile Function: A multidimensional scale for assessment of erectile dysfunction. *Urology* **49**, 822-830 (1997).
21. O'Leary, M. et al. A brief male sexual function inventory for urology. *Urology* **46**, 697-706 (1995).
22. Derogatis, L.R. & Melisaratos, N. The DSFI: a multi-dimensional measure of sexual functioning. *Journal of Sex and Marital Therapy* **5**, 244-281 (1979).
23. Libman, E., Rothenberg, I., Fichten, C.S. & Amsel, R. The SSES-E: a measure of sexual self-efficacy in erectile functioning. *Journal of Sex and Marital Therapy* **11**, 233-244 (1985).
24. Fikree, F.F., Gray, R.H. & Shah, F. Can men be trusted? A comparison of pregnancy histories reported by husbands and wives. *American Journal of Epidemiology* **138**, 237-242 (1993).
25. Cappelleri, J.C., Rosen, R.C., Smith, M.D., Mishra, A. & Osterloh, I.H. Diagnostic evaluation of the erectile function domain of the International Index of Erectile Function. *Urology* **54**, 346-351 (1999).
26. StataCorp. Stata Statistical Software. Release 7.0 edn (Stata Corporation, College Station, TX., 2001).

This page has been left blank intentionally
for formatting purposes

15 Mental Health

Chapter summary

Mood disorders (depression and anxiety) were assessed using a variety of methods: a computerised assessment program (Composite International Diagnostic Interview, or CIDI) administered by a psychologist; validated self-completed questionnaires (Kessler 10-item and General Health Questionnaire 12-item); and self-reported diagnoses of depression, anxiety or other somatic symptoms, and current medications (from the Postal Questionnaire). There was a high level of agreement between the self-reported indicators of mood disturbances and the objective tests administered during the health examination. The exposed group was more likely to report a previous diagnosis of depression and/or anxiety, had higher use of anti-depressant medications, and had nearly a two-fold increased risk of diagnosis of depression and anxiety using the CIDI. Results were strong and consistent in that they were significant in the overall analysis, in both Program 1 and 2 subgroup analyses, and showed a dose-response relationship. Data from other validated questionnaires also indicated that the exposed group were more likely to have mental distress and social dysfunction than both comparison groups and than the Australian population in general.

Chapter contents

15.1 Introduction

In his report to the Board of Inquiry (BOI), Dr Eric Donaldson presented data from 105 Health Questionnaires completed by individuals involved in DSRS activities. A total of 47% of individuals reported some neurological/psychological symptoms.[1] Symptoms listed by the respondents included anxiety/stress, claustrophobia, depression, indecisiveness, irritability, mood swings, paranoia, loss of memory/lapses, psychological problems, exhaustion, headaches, head pain. This information, together with the number of F-111 DSRS workers who registered problems with mood and depression with the DVA F-111 Interim Health Care Scheme, was considered sufficient to warrant further investigation.

The issue of mood and depressive disorders following occupational exposure has been debated in the literature for many years. The historical background for a possible link between organic solvents and mental health was described by Arlien-Solbørg,[2] dating back to 1863. Delpech described acute and sub-chronic intoxication in workers exposed to carbon disulphide including effects on mood, insomnia and memory problems.[2] Reports of psychiatric symptoms associated with occupational exposure to organic solvents have appeared regularly since the 1970s.[3-9] In fact, alterations in mood and anxiety are reported to be the most common finding for solvent-exposed persons.[10] Further, it has been proposed that psychological distress (i.e. personality changes and depression) may be the earliest manifestations of exposure.[11] A number of studies – including White et al.,[12] Parkinson,[13] Morrow et al.,[14] Morrow et al.,[15] Stollery and Flindt[16] – indicated clinically significant levels of depression, anxiety, somatic concerns and disturbances in thinking, in solvent-exposed individuals compared to comparisons. This association has also been documented in military populations[17] including munitions plant workers.[18] These effects may be mediated biologically or may be the result of psychological stressors and sociocultural factors,[19] e.g. situations of low control, traumatic working conditions. Even in situations of definite central nervous system (CNS) toxicity, psychological and emotional reactions may play a part in the persistence of symptoms and may complicate recovery and exacerbate disability. In one study[20] high levels of emotional distress at the time of the original exposure were found to be a risk factor for persistent neuropsychological deficits and disability.

Also of concern were neurasthenic symptoms such as persistent feelings of fatigue after quite minor mental or physical effort, possibly also involving dizziness, tension headaches, sleep problems, anxiety, inability to relax, and irritability.[21] Fidler et al.[22] reported a positive association between increasing exposure and neurasthenic symptoms.

15.2 Measures

15.2.1 Health examination

Five different measures of mood disorder were used during the SHOAMP health examination:

1) the K-10 (10-item Kessler Psychological Distress Scale)[23] – as a measure of anxiety and depressive symptoms
2) the 12-item General Health Questionnaire (GHQ-12) – as a screening tool to identify short-term changes in mental health[24]
3) the computer-administered Composite International Diagnostic Interview (CIDI)[25] version 2.1 – to collect data about depression
4) the CIDI, again – to collect data about anxiety
5) computerised neurasthenia* items from the National Mental Health Interview program, administered by the attending HSA psychologist.

Kessler Psychological Distress Scale (K-10)

The K-10 was developed to screen populations on psychological distress, specifically to measure symptoms of anxiety and depression over the four weeks prior to the test. The scale has been used in a number of population health surveys in Australia, such as State-specific population surveys and in the National Survey of Mental Health and Wellbeing.[26-30] The K-10 has been found to be both valid and reliable, and its brevity and simplicity made it particularly appropriate for use in SHOAMP. It was administered to SHOAMP participants during the five minutes between their first and second seated blood pressure measures; though this was not optimal timing, it was done to expedite an already very long examination. Because the same procedure was carried out equally with both exposed and comparison participants, it was not considered a source of bias. The K-10 was scored using the following method:

* Neurasthenia is defined by the World Health Organisation International Classification of Disease and Related Disorders (ICD-10) as either complaints of fatigue after minor mental effort, or complaints of fatigue and weakness after minor physical effort, lasting at least three months. The person is unable to recover by means of rest, relaxation or entertainment. These indications are accompanied by a variety of other physical symptoms such as muscular aches and pains, dizziness, tension headaches, sleep disturbance, inability to relax, and irritability. The disorder does not occur in the presence of a depressive episode, hypomania, mania, bipolar affective disorder, panic disorder, or generalised anxiety disorder.

1) Values were reversed for each category with 5 = ”all of the time” and 1 = ”none of the time”. These items were then summed to provide scores ranging from 10 to 50, with a higher score indicating higher probability of having anxiety or depression.

2) In order to give some clinical context to these numbers, the Clinical Research Unit for Anxiety and Depression at the University of NSW developed cut-off scores for the K-10: a score of 10-15 meant low or no probability of having anxiety or depression; 16-29 meant medium probability; and a score of 30-50 meant high probability. It was this scale which has been used to analyse results in the current study.31

General Health Questionnaire (GHQ-12)

The 12-item General Health Questionnaire (GHQ-12) is a screening test designed to identify short-term changes in mental health: depression, anxiety, social dysfunction, and somatic symptoms.[32] The GHQ-12 has high levels of “test – re-test” reliability, sensitivity and specificity[33] and has generally been accepted as a valid screening instrument.[34] The questionnaire comprises twelve questions, asking each person about their general level of happiness, experience of depressive and anxiety symptoms, and sleep disturbance over the last four weeks. Attending nurses from HSA asked study participants to rate the degree to which they had experienced each symptom, using a four-point Likert scale. Three scoring methods have been reported, for all of which low scores indicate low probability of having mental illness.[16] The standard method scores symptomatic responses (e.g. “more than usual” and “much more than usual”) as “1” and non-symptomatic responses (“not at all”, “no more than usual”) as “0”, and sums over the 12 items. Another method assigns scores of 0-3 to each response and sums over items, resulting in a score ranging from 0 to 36. However, within an Australian setting, the “C-GHQ” scoring method is preferable.[35] This method is designed to increase the likelihood of detecting chronic disorders by scoring the “no more than usual” response to negative items as “1”. Positive items are scored using the standard binary method.

Threshold scores indicating the likelihood of mental illness have been empirically determined in primary care and community settings and in several countries. Donath[36] found that a threshold of 3/4 gave the best sensitivity for a given specificity using data from the Australian 1997 National Mental Health Survey. Thus the SHOAMP data was scored with the C-GHQ method, and respondents with scores of four and over were classified as having a high probability of mental illness.

Composite International Diagnostic Interview (CIDI)

The CIDI[25] is a highly-structured, fully-standardised interview which assesses mental disorders according to the diagnostic criteria specified by DSM-IV (Diagnostic and Statistical Manual of Mental Disorders) and ICD-10 (International Statistical Classification of Diseases and Related Health Problems). The CIDI covers eight domains of mental health: somatoform disorders, anxiety disorders, depressive disorders, mania, schizophrenia, eating disorders, cognitive impairment, and substance use disorders. It is designed for use by non-clinicians after training. While its reliability has been found to be satisfactory, the validity of the CIDI is incompletely established,[37,38] particularly against the criterion of a standardised assessment by a psychiatrist experienced in the syndrome under consideration. However, the CIDI does provide a standardised and reproducible assessment when in-depth interviews are not feasible. For SHOAMP, an attending HSA psychologist administered two computerised modules from the CIDI-Auto which explored the issues of anxiety and depression separately. For SHOAMP, the CIDI software was specially configured so that only the anxiety and depression modules were available to HSA psychological staff to administer. The CIDI also included questions on the timing of symptoms, including date of first episode and most recent episode.

Instructions were provided for saving the electronic files from each participant interview and for printing the DSM-IV descriptors for each participant as to whether they met diagnostic criteria for anxiety and/or depressive disorder (according to any of the ICD-10 or DMS-IV diagnoses). General demographic data were entered for each study participant: SHOAMP 7-digit ID code, gender, age, and date of birth. The program then automatically stepped through a series of screening items to identify where anxious or depressive behaviours were being experienced.

To ensure data quality during the administration of the CIDI-Auto, attending psychologists were given the following key directives:

- Follow the rules for entering data, as described in the Administrative Guide and Reference manual.
- Read questions exactly as they are worded.
- Read the entire question; finish or repeat the question if a respondent answers prematurely.
- Emphasise key words and time frames.
- Follow rules for clarification and probing, as described in the Administrative Guide and Reference manual.

- Follow rules for giving feedback, as described in the Administrative Guide and Reference manual.[25]

In the context of the SHOAMP General Health and Medical Study, it was not appropriate to make clinical interpretations or to offer counselling. Follow-up with the participant's preferred medical practitioner was encouraged where referral was necessary. The collected CIDI data were used descriptively to report results to participants to describe whether they had, or had not, met the diagnostic criteria for anxiety and/or depression. These data were converted to a format designed to be used with statistical software packages.

Each CIDI test consisted of 4 file types: ".out", ".scs", ".ini" and ".all". The .scs file gave the results of the test, in a text format according to DSM-IV and ICD-10 diagnoses. One file set was stored for each participant. In order to provide useable data, each file set required re-scoring. Once all the results had been collected and checked, all the .out, .ini and .all files were gathered and, using the scoring program provided with the CIDI program, every test was re-scored in one large batch to give a data, rather than a text, output.

Neurasthenia

Computerised neurasthenia items were administered by the attending HSA psychologist. The successful completion of a neurasthenia interview resulted in the automatic creation of 5 output files. Each file was recognised by a different extension: ".out", ".scs", ".ini", ".all" and ".dat". Once all the health examination results were available, all the files were re-scored using the same scoring program as that used in the National Survey of Mental Health and Well-Being.

15.2.2 Postal Questionnaire

In addition to clinical assessment of depression and anxiety, Postal Questionnaire items 2.11 and 2.12 asked if the participant had ever been previously diagnosed by a physician as suffering from, respectively, "anxiety" and/or "depression". For neurasthenia, Postal Questionnaire item 2.10 asked each person to nominate yes or no to experiencing "fatigue" in the past month. Two medication outcomes were also analysed: the use of anxiolytic medications (ATC code N058) and anti-depressants (ATC code N06A).

15.3 Potential confounders

Potential confounders for mood disorder outcomes included age (categorised into 5-year intervals), posting, rank, HSA centre, alcohol intake status, smoking status, education status, civilian exposure to organic solvents, and responses to Postal Questionnaire item3.1, “has anyone in your immediate family ever suffered from depression”. For Postal Questionnaire items, HSA centre was not included as a potential confounder.

15.4 Analyses

All analyses followed the outline described in Chapter 6 (Analysis). All subjects were screened using the Rey-15 item test,[39] included in the SHOAMP health examination battery to examine the potential for inadequate effort during psychological testing. Scores of eight or less on this test were judged to be an indication of potentially unreliable results on the other mood and cognition tests; thus individuals with scores within this range were excluded from further analyses.[40,41]

Each response to the Kessler 10-item scale was reversed and then the scale was summed. Following summation, the scale was categorised into three levels of probability of having anxiety or depression: low probability for those scoring less than or equal to 15, medium probability for those scoring 16 to 29, and high probability for those scoring 30 or more. Further categorisation into two groups was then undertaken (collapsing medium and high probability groups), and analysis was by multiple logistic regression.

The distribution of the outcome (C-GHQ score greater than 4) was examined in the exposure and comparison groups. After ensuring that the relationship between exposure group and outcomes across HSA centre was homogeneous, multiple logistic regression was conducted using exposure group and potential confounders as explanatory variables. The multiple logistic regression analyses were also conducted separately for those whose exposure related to Program 1 and Program 2 exposures and for dose-response.

From the CIDI, all depression codes and all anxiety codes were collapsed into the two categories of depression or anxiety. In all three study groups, those with date of onset and date of most recent episode occurring before the start of their posting period were excluded, since their disorders would not be related to DSRS activities. Duplicates were removed and

participants were classified as having a depressive and/or anxiety disorder or not. Depression and anxiety were analysed separately using multiple logistic regression.

Participants were classified as having neurasthenia or not. All those classified with neurasthenia with the last date of onset prior to their posting date were excluded. Analysis was by multiple logistic regression.

The distributions of self-reported data on anxiolytic (coded as NO5B) and antidepressant (NO6A) medications were examined in the exposure and comparison groups for depression and anxiety. Multiple logistic regression was conducted using exposure group and potential confounders as explanatory variables. The multiple logistic regression analyses were also conducted separately for those whose exposure related to Program 1 and Program 2 exposures and for dose-response.

15.5 Results

15.5.1 Kessler 10-item scale

Of the 1538 participating in the health examination and eligible for analysis, there were 17 who did not successfully complete the Rey 15-item test. Additionally, three participants who passed the Rey did not complete the Kessler; one participant refused both the Rey and the Kessler; another refused the Rey but not the Kessler; two participants did not complete the Rey but completed the Kessler; and two participants did not complete either the Rey or the Kessler. Consequently there were 1512 records available for analysis. There were 80 participants classified as having a high probability of anxiety or depression (Table 15.1), with 6% from Amberley, 2% from Richmond, and 7% from the exposed. These were too few to analyse, so high and medium probability categories were combined, resulting in 32% from Amberley, 29% from Richmond and 50% from exposed (see Table 15.2). There was no evidence of heterogeneity by the Breslow-Day test for Amberley versus exposed (p=0.96) or for Richmond versus exposed (p=0.61).

Table 15.1 : Classification of responses to the Kessler scale into the three probability categories before collapsing for analysis

Kessler 10-item scale	Group			Total
	Amberley	Richmond	Exposed	
Low probability of having anxiety or depression	274 68.50	363 71.18	303 50.33	940
Medium probability of having anxiety or depression	102 25.50	135 26.47	255 42.36	492
High probability of having anxiety or depression	24 6.00	12 2.35	44 7.31	80
Total	400	510	602	1512

Frequency missing = 3

Table 15.2 : Kessler scale – Summary of responses following the combining of high and medium probability categories

Kessler 10-item scale	Group			Total
	Amberley	Richmond	Exposed	
Medium/High probability of having anxiety or depression	126 31.50	147 28.82	299 49.67	572
Low probability of having anxiety or depression	274 68.50	363 71.18	303 50.33	940
Total	400	510	602	1512

There was a significant association between group and Kessler category of medium to high probability of having anxiety or depression when accounting for all potential confounders for all exposed and Program 1 and Program 2 subgroups ($p<0.0001$) (Table 15.3). This was consistent in the reduced model and with the robust standard variance estimates. Specifically, for all exposed, the odds of being classified as medium/high probability of having anxiety or depression in the Amberley group were 0.37 (0.26, 0.51) times those for the exposed group, and the odds of being classified as medium/high probability for the

Richmond group were 0.41 (0.31, 0.55) times those of the exposed group. Similar odds were observed across programs for Amberley (0.32, Program 1 and 0.37, Program 2) and Richmond (0.37, Program 1 and 0.43, Program 2). There was a significant association between dose and medium/high classification (p<0.0001) but no clear trend of increase with increasing dose, where the odds of medium/high probability for the mild exposure group compared to the unexposed were 1.65 (1.12, 2.44) , for the moderate exposure group 2.91 (2.0, 4.23), and for the prolonged exposure group 2.84 (1.96, 4.12).

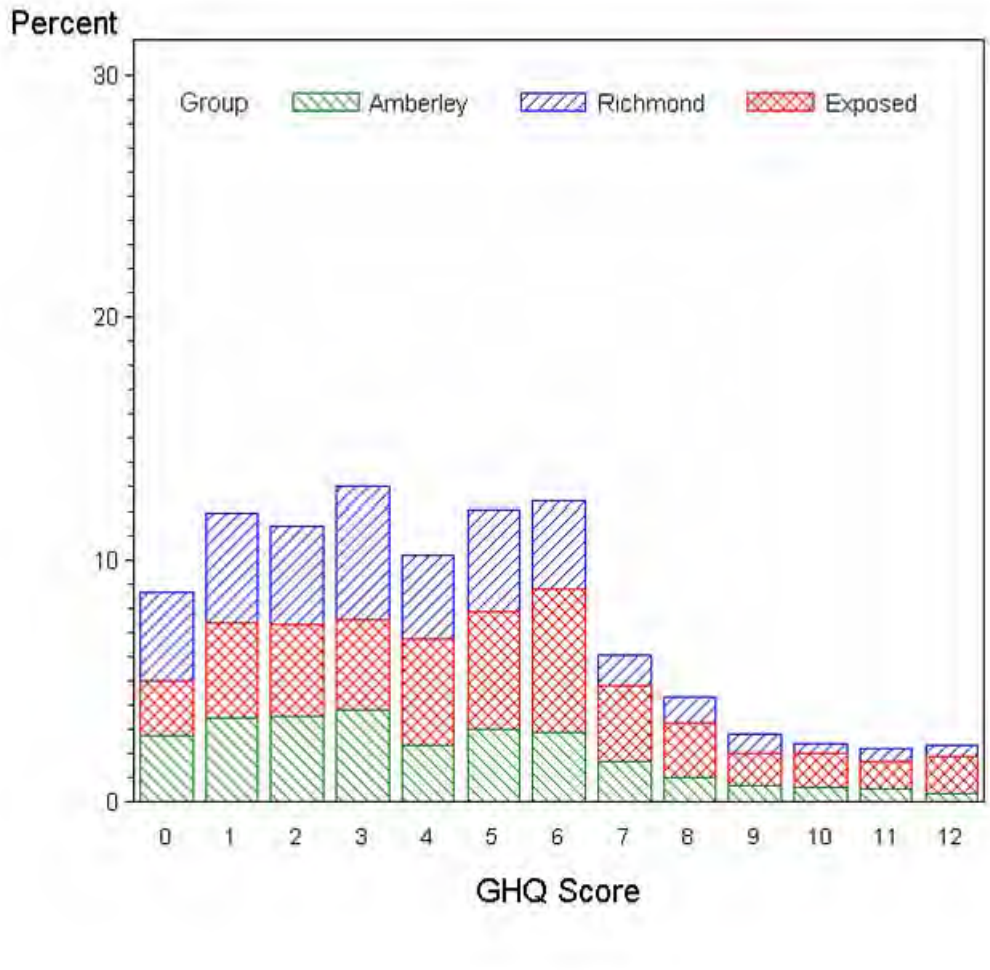
Table 15.3 : Kessler scale – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.37	0.26	0.51	2	49.31	<.0001
	Richmond vs Exposed	0.41	0.31	0.55	.	.	.
Program 1	Amberley vs Exposed	0.32	0.21	0.49	2	35.82	<.0001
	Richmond vs Exposed	0.37	0.25	0.53	.	.	.
Program 2	Amberley vs Exposed	0.37	0.24	0.58	2	22.62	<.0001
	Richmond vs Exposed	0.43	0.29	0.65	.	.	.
Dose	Mild exposure vs Unexposed	1.65	1.12	2.44	3	48.55	<.0001
	Moderate exposure vs Unexposed	2.91	2.00	4.23	.	.	.
	Prolonged exposure vs Unexposed	2.84	1.96	4.12	.	.	.

15.5.2 General Health Questionnaire

Of the 1538 participants who took part in the health examination, five did not complete the GHQ-12 and nine missed at least one of the items, which meant that an overall score could not be summed (eight from the exposed group, and three from each comparison groups). A further 21 failed or did not complete the REY-15-item test and were excluded. Therefore, 1503 were included in the analysis, of which 20 were female. There was no heterogeneity between offices for the Amberley comparison group (p=0.40) or the Richmond comparison group (p=0.30). The distribution of C-GHQ scores was strongly skewed (meaning fewer participants had higher scores) (see Figure 15.1).

Figure 15.1 : Frequency distribution of GHQ scores for exposed and comparison groups



Mean and median scores were higher in the exposed group. Interestingly, the mean scores for all groups were close to or higher than four, the threshold for high probability of having mental illness. Numbers and percentages of those falling above and below the C-GHQ threshold in each group are shown in Table 15.4. The proportion classified as having high probability of mental illness was higher in the exposed group than in the other groups.

Table 15.4 : Number and percentage of participants classified as being at low and high probability of mental illness by the GHQ-12

	Group			
Result from GHQ-12	**Amberley**	**Richmond**	**Exposed**	**Total**
Low probability < 4	203 51.01	269 52.85	206 34.56	678
High probability ≥ 4	195 48.99	240 47.15	390 65.44	825
Total	398	509	596	1503

Frequency Missing =12

The multiple logistic regression analyses, modelling on the category of having a high probability of anxiety or depression, are summarised in Table 15.5. The odds of being in the high probability category were statistically significantly lower for the comparison groups (0.36 for Amberley and 0.48 for Richmond) and this was consistent in Program 1 and Program 2 subgroups. The results were similar in the reduced model, and the confidence intervals were reduced with the robust standard variance estimates. As shown in Appendix 15G, being in the oldest age group, being a current smoker, and having a high school education (as compared to trade qualifications), increased the odds of being at a higher chance of having mental illness. The odds of being in the high category increased with increasing exposure from 1.64 to 2.45 to 2.93 for mild, moderate and prolonged exposure to DSRS activities.

Table 15.5 : GHQ – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.36	0.26	0.50	2	45.29	<.0001
	Richmond vs Exposed	0.48	0.36	0.63			
Program 1	Amberley vs Exposed	0.38	0.25	0.57	2	24.44	<.0001
	Richmond vs Exposed	0.47	0.33	0.68			
Program 2	Amberley vs Exposed	0.32	0.21	0.50	2	27.53	<.0001
	Richmond vs Exposed	0.44	0.30	0.65			
Dose	Mild exposure vs Unexposed	1.64	1.12	2.38	3	42.64	<.0001
	Moderate exposure vs Unexposed	2.45	1.67	3.58			
	Prolonged exposure vs Unexposed	2.93	2.00	4.31			

15.5.3 Self-reported physician-diagnosed depression and anxiety

Of the 1726 participants returning a Postal Questionnaire, 41 did not complete the question relating to physician-diagnosed depression. Consequently, there were 1685 records available for analysis. Of these, 29% of the exposed, 19% of Amberley participants and 16% of Richmond participants reported a physician diagnosis of depression (Table 15.6).

Table 15.6 : PQ item self-reported physician diagnosed depression – Unadjusted proportions of positive responses from the different groups

Self-reported physician diagnosed depression	Group			Total
	Amberley	Richmond	Exposed	
Depression	91 19.00	93 15.98	182 29.17	366
No Depression	388 81.00	489 84.02	442 70.83	1319
Total	479	582	624	1685

Frequency missing = 31

When considering all potential confounders in multiple logistic regression, there was significant association between self-reported physician-diagnosed depression and group for all exposed ($p<0.0001$), Program 1 ($p<0.0001$) and Program 2 ($p=0.0016$). Specifically, for all exposed, the odds of reporting physician diagnosed depression and being in the Amberley group were 0.49 (0.34, 0.70) times those of the exposed, and the odds for Richmond were 0.45 (0.33, 0.62) times those of the exposed (see Table 15.7). These odds varied little for Program 1, Program 2, or the reduced model. There was a significant association between exposure dose and self-reported physician-diagnosed depression ($p<0.0001$), with an increasing trend where the odds of being in the mild exposed group and self-reporting a physician diagnosis of depression were 1.80 (1.18, 2.74) times those of the unexposed, 1.83 (1.22, 2.74) for moderate exposure, and 2.75 (1.89, 4.0) for prolonged exposure.

Of the 1726 participants returning a Postal Questionnaire, 45 did not complete the question relating to physician-diagnosed anxiety. Consequently, there were 1681 records available for analysis. Of these, 22% of the exposed, 13% of Amberley participants and 12% of Richmond participants reported a physician diagnosis of anxiety (Table 15.8).

Table 15.7 : PQ item self-reported physician diagnosed depression – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.49	0.34	0.70	2	30.17	<.0001
	Richmond vs Exposed	0.45	0.33	0.62			
Program 1	Amberley vs Exposed	0.46	0.30	0.70	2	22.93	<.0001
	Richmond vs Exposed	0.42	0.29	0.61			
Program 2	Amberley vs Exposed	0.49	0.31	0.78	2	12.92	0.0016
	Richmond vs Exposed	0.49	0.32	0.74			
Dose	Mild exposure vs Unexposed	1.80	1.18	2.74	3	31.84	<.0001
	Moderate exposure vs Unexposed	1.83	1.22	2.74			
	Prolonged exposure vs Unexposed	2.75	1.89	4.00			

Table 15.8 : PQ item self-reported physician diagnosed anxiety – Unadjusted proportions of positive responses from the different groups

Self-reported physician diagnosed anxiety	Group			Total
	Amberley	Richmond	Exposed	
Anxiety	64 13.45	69 11.90	135 21.60	268
No anxiety	412 86.55	511 88.10	490 78.40	1413
Total	476	580	625	1681

Frequency missing = 35

When considering all potential confounders in multiple logistic regression, there was significant association between self-reported physician-diagnosed anxiety and group for all exposed (p=0.0001), Program 1 (p<0.0001) and Program 2 (p=0.015). Specifically, for all exposed, the odds of reporting physician-diagnosed anxiety and being in the Amberley group were 0.55 (0.37, 0.83) times those of the exposed, and the odds for Richmond were 0.49 (0.35, 0.70) times those of the exposed (see Table 15.9). These odds varied little for Program 1, Program 2, or the reduced model. There was a significant association between exposure dose and self-reported physician-diagnosed anxiety (p<0.0001) but no clear stepwise increase in risk with increasing dose; the odds of being in the mild exposed group and self-reporting a physician diagnosis of anxiety were 1.95 (1.24, 3.07) times those of the unexposed, 1.57 (1.0, 2.47) for moderate exposure, and 2.46 (1.63, 3.72) for prolonged exposure.

Table 15.9 : PQ item self-reported physician diagnosed anxiety – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.55	0.37	0.83	2	18.43	<.0001
	Richmond vs Exposed	0.49	0.35	0.70			
Program 1	Amberley vs Exposed	0.48	0.30	0.77	2	19.32	<.0001
	Richmond vs Exposed	0.42	0.28	0.63			
Program 2	Amberley vs Exposed	0.55	0.33	0.94	2	8.35	0.0154
	Richmond vs Exposed	0.51	0.32	0.81			
Dose	Mild exposure vs Unexposed	1.95	1.24	3.07	3	21.93	<.0001
	Moderate exposure vs Unexposed	1.57	1.00	2.47			
	Prolonged exposure vs Unexposed	2.46	1.63	3.72			

15.5.4 CIDI diagnoses of depression and anxiety

Of the 1538 eligible participants in the health examination, 35 who passed the Rey and six who refused or otherwise did not complete the Rey, did not undertake the CIDI. Additionally, 17 participants did not successfully complete the Rey test. Consequently there were 1480 records available for analysis. There was no evidence of heterogeneity with the CIDI for depression (Amberley versus exposed, p=0.66; Richmond versus exposed, p=0.34), but there was slight indication for anxiety (Amberley versus exposed, p=0.02; Richmond versus

exposed, p=0.06). It was considered reasonable to assume that this positive test for heterogeneity was a statistical artifact for the following reasons:

- the large power of the Breslow-Day chi-square test for homogeneity of odds ratio
- there was only one significance value out of four significance tests
- the CIDI module was applied for depression and anxiety at the same time
- when examining the pattern of associations within offices, only two of the eight offices went against the trend of higher proportion of exposed classified with anxiety.

In regard to depressive disorder, approximately 30% of the exposed classified as having had depression, compared with 22% from Amberley and 17% from Richmond (see Table 15.10). Similarly, 26% of the exposed classified as having experienced anxiety, compared with 17% from Amberley and 11% from Richmond (see Table 15.11). When all potential confounders were considered in multiple logistic regression, group was significantly associated with both depression and anxiety for all exposed (p=0.0003, p<0.0001 respectively) and for Program 1 (p=0.0003, p=0.0001 respectively), but only for anxiety for Program 2 (p=0.05, p=0.002 respectively) (see Table 15.12).

Table 15.10 : Depressive disorder – Unadjusted proportions of positive responses from the different groups

	Group			
CIDI Depression	**Amberley**	**Richmond**	**Exposed**	**Total**
Diagnostic criteria for depression met	85 22.02	87 17.19	177 30.10	349
Diagnostic criteria for depression not met	301 77.98	419 82.81	411 69.90	1131
Total	386	506	588	1480

Frequency missing = 35

Table 15.11 : Anxiety disorder – Unadjusted proportions of positive responses from the different groups

CIDI anxiety	Group			Total
	Amberley	Richmond	Exposed	
Diagnostic criteria for anxiety met	64 16.58	58 11.46	151 25.68	273
Diagnostic criteria for anxiety not met	322 83.42	448 88.54	437 74.32	1207
Total	386	506	588	1480

In particular, the odds of being diagnosed by the CIDI as having been or being depressed and in the Amberley group were 0.59 (0.41, 0.85) times those of the exposed for all exposed and for Richmond were 0.53 (0.38, 0.74) for all exposed. There were similar odds for both Programs 1 and 2; however, the 95% confidence intervals for the Amberley odds in Program 2 contained one. There was a significant association of CIDI diagnosis of depression and dose with a stepwise progression from 1.32 (0.85, 2.06) for mild exposure, 1.60 (1.05, 2.44) for moderate exposure, and 2.45 (1.65, 3.66) for prolonged exposure. The 95% confidence intervals for mild exposure contained one and were not considered significant. Results were similar for the reduced model and for the robust standard variances.

Table 15.12 : Depression disorder – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.59	0.41	0.85	2	15.95	0.0003
	Richmond vs Exposed	0.53	0.38	0.74			
Program 1	Amberley vs Exposed	0.54	0.35	0.85	2	16.06	0.0003
	Richmond vs Exposed	0.44	0.30	0.66			
Program 2	Amberley vs Exposed	0.66	0.41	1.07	2	5.97	0.0504
	Richmond vs Exposed	0.58	0.37	0.90			
Dose	Mild exposure vs Unexposed	1.32	0.85	2.06	3	20.30	0.0001
	Moderate exposure vs Unexposed	1.60	1.05	2.44			
	Prolonged exposure vs Unexposed	2.45	1.65	3.66			

The odds of being diagnosed by the CIDI as having had anxiety and in the Amberley group were 0.63 (0.42, 0.95) times those of the exposed for all exposed, and for Richmond were 0.44 (0.30, 0.64) for all exposed (see Table 15.13). There were similar odds for both Programs 1 and 2. There was a significant association of CIDI diagnosis of having had an anxiety disorder and dose (p=0.0001) but with no clear stepwise progression from 1.78 (1.11, 2.85) for mild exposure, 1.61 (1.02, 2.55) for moderate exposure and 2.62 (1.71, 4.03) for prolonged exposure. Inferences remained the same with the reduced model, but the comparison with Amberley lost significance when using the robust standard variances.

Table 15.13 : CIDI Anxiety – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.63	0.42	0.95	2	18.51	<.0001
	Richmond vs Exposed	0.44	0.30	0.64			
Program 1	Amberley vs Exposed	0.53	0.33	0.86	2	18.13	0.0001
	Richmond vs Exposed	0.38	0.25	0.60			
Program 2	Amberley vs Exposed	0.55	0.32	0.95	2	12.50	0.0019
	Richmond vs Exposed	0.40	0.25	0.67			
Dose	Mild exposure vs Unexposed	1.78	1.11	2.85	3	20.73	0.0001
	Moderate exposure vs Unexposed	1.61	1.02	2.55			
	Prolonged exposure vs Unexposed	2.62	1.71	4.03			

15.5.5 Neurasthenia

There were 21 participants in the health examination who did not undertake the neurasthenia examination, and six of these did not complete the Rey. A further 10 participants, including one who did not satisfactorily complete the Rey, did not have age at onset recorded, and so were excluded from further analysis. A further 16 were excluded based on their performance on the Rey. This left 1491 available for analysis. There was no evidence of heterogeneity with the Breslow-Day test for homogeneity for Amberley versus exposed (p=0.16), nor for

Richmond versus exposed (p=0.83). Approximately 10% of exposed, 7% of Amberley and 4% of Richmond were classified as having had neurasthenia (Table 15.14).

Table 15.14 : Neurasthenia – Unadjusted proportions of positive responses from the different groups

	Group			
Neurasthenia	**Amberley**	**Richmond**	**Exposed**	**Total**
Present	27 6.89	19 3.75	58 9.63	104
Absent	365 93.11	487 96.25	544 90.37	1396
Total	392	506	602	1500

Frequency missing = 15

Neurasthenia was significantly associated with group for all exposed (p=0.016) and for the Program 1 subgroup (p=0.04) but not for the Program 2 subgroup(p=0.14). The odds ratio estimates were all below one for all exposed, Program 1 and Program 2; however, the 95% confidence intervals spanned one for Amberley in all exposed and Program 1, and for both Amberley and Richmond in Program 2. Dose was almost significantly associated with neurasthenia (p=0.07), but there was no clear step trend of increase with increasing dose going from 1.23 (0.57, 2.62) for mild exposure to 2.08 (1.10, 3.94) for moderate exposure and 1.97(1.02, 3.80) for prolonged exposure (see Table 15.15). The results were similar for the reduced model and for the robust variance estimates.

Table 15.15 : Neurasthenia – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.72	0.40	1.31	2	8.28	0.0159
	Richmond vs Exposed	0.41	0.23	0.76			
Program 1	Amberley vs Exposed	0.70	0.34	1.46	2	6.57	0.0374
	Richmond vs Exposed	0.40	0.19	0.81			
Program 2	Amberley vs Exposed	0.87	0.39	1.93	2	3.99	0.1362
	Richmond vs Exposed	0.48	0.22	1.07			
Dose	Mild exposure vs Unexposed	1.23	0.57	2.62	3	6.92	0.0746
	Moderate exposure vs Unexposed	2.08	1.10	3.94			
	Prolonged exposure vs Unexposed	1.97	1.02	3.80			

15.5.6 Medications

This analysis included data on all those participants who completed the Postal Questionnaire (n=1716). Table 15.16 shows the numbers and percentages of participants who were taking at least one type of anti-depressant. The multiple logistic regression analyses for depression medications, modelling on the “taking medication” category, are summarised in Table 15.17. The odds of being on anti-depressants were lower for both the Amberley and Richmond comparisons than for the exposed group, with odds ratios of 0.63 and 0.51 respectively. Although the Amberley comparison was not statistically significant in the full model, it was significant in the reduced model. Results were consistent in Program 1 and 2 subgroups, but there was no clear dose-response effect. Numbers in the anxiolytic group were too small for multivariate analyses (Table 15.18).

Table 15.16 : Anti-depressant medications – Unadjusted proportions of positive responses from the different groups

Self-reported anti-depressant medications	Group			Total
	Amberley	Richmond	Exposed	
No reported anti-depressant meds	454 92.84	559 94.43	566 89.13	1579
Reported anti-depressant meds	35 7.16	33 5.57	69 10.87	137
Total	489	592	635	1716

Table 15.17 : Anti-depressant medications – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.63	0.37	1.06	2	8.17	0.0168
	Richmond vs Exposed	0.51	0.32	0.83	.	.	.
Program 1	Amberley vs Exposed	0.54	0.29	0.98	2	9.69	0.0079
	Richmond vs Exposed	0.43	0.25	0.74	.	.	.
Program 2	Amberley vs Exposed	0.72	0.35	1.45	2	4.25	0.1197
	Richmond vs Exposed	0.51	0.27	0.98	.	.	.
Dose	Mild exposure vs Unexposed	0.75	0.34	1.66	3	21.88	<.0001
	Moderate exposure vs Unexposed	3.03	1.81	5.09	.	.	.
	Prolonged exposure vs Unexposed	1.89	1.07	3.35	.	.	.

Table 15.18 : Anxiolytic medications – Unadjusted proportions of positive responses from the different groups

Self-reported anti-anxiety medications	Group			Total
	Amberley	Richmond	Exposed	
No reported anti-anxiety meds	475 97.14	589 99.49	622 97.95	1686
Reported anti-anxiety meds	14 2.86	3 0.51	13 2.05	30
Total	489	592	635	1716

15.6 Discussion

The two major indicators of mental disorder in this study were anxiety and depression, each measured by computerised interview (CIDI), self-report, and medication use. In addition, two other validated scales, the K-10 and GHQ-12, were used to gauge combined anxiety and depression and provide a measure of overall mental health. Moreover, the neurasthenia module was used as a measure of overall non-specific psychological distress.

Depression

Self-report of previous physician-diagnosed depression was higher in the exposed group than the comparison groups, i.e. 29% versus 19% in Amberley and 16% in Richmond. This translates into a roughly two-fold increase in the diagnosis of depression in the F-111 DSRS group. This is significant in the overall analysis and in the Program 1 and 2 subgroup analyses, and there is a dose-response effect.

This result was generally congruent with medication use. The proportion of the exposed group taking at least one anti-depressant was also elevated: 11% versus 7% in the Amberley and 6% in the Richmond groups. This leads to an odds ratio of 1.7, which was significant in the overall and subgroup analyses, mainly against the Richmond comparisons, but not against the Amberley comparisons. Although the global test for dose-response was significant, there was not a stepwise increment in risk estimates with increasing exposure.

Results from the CIDI Depression module were consistent with the previous results. The CIDI diagnosed 29% of the exposed group as having been depressed compared with 21% of the Amberley group and 17% of the Richmond group. This equates to a 1.8-fold increased probability of having depression; this is significant in the overall analysis and in both Program 1 and 2 subgroups analyses, and shows a dose-response effect. The prevalence of depression may appear high in all three groups compared to its prevalence in the Australian population – 4.2% in males in the National Survey of Mental Health (NSMH)[42] – but this probably relates to the timeframe of questioning. The NSMH focused on depression occurring within the last 12 months, whereas the CIDI detects depression at any time in the past, including that which occurred previous to any DSRS activities. We were therefore careful to exclude diagnoses of depression which started and ended before the exposure period of interest.

Overall, there is noteworthy concordance between self-reported physician diagnosis of depression, use of anti-depressants, and depression as judged by the CIDI module; and it shows a strong and convincing association in this area.

Anxiety

Self-report of physician-diagnosed anxiety was higher in the exposed group than in the comparison groups, i.e. 22% versus 13% in Amberley and 12% in Richmond. This translates into a roughly two-fold increase in the diagnosis of anxiety in the DSRS group. This is significant in the overall analysis and in the Program 1 and 2 subgroup analyses, and there is the suggestion of a dose-response effect.

Too few subjects were taking anxiolytics (anti-anxiety medications) to do any meaningful analysis (2% in the exposed group, 2.8% in Amberley, and 0.5% in Richmond). However, results from the CIDI Anxiety module were consistent with self-report results. The CIDI diagnosed 25% of the exposed group with an anxiety disorder, versus 16% of the Amberley group and 11% of the Richmond group. This translates into an odds ratio of approximately 1.9. This increased probability of having an anxiety disorder is significant in the overall analysis and in both Program 1 and 2 subgroups analyses, and suggests a dose-response effect.

Again, there is a high degree of concordance between self-reported anxiety, medication use, and CIDI diagnosis of an anxiety disorder, and this indicates a strong and convincing association.

Combined anxiety/depression and general mental health

The Kessler 10-item scale (K-10) is a screening measure for anxiety and depression occurring within the last four weeks. The following clinical cut-off values have been established in the Australian population:

- *0-15: low probability of having anxiety or depression.* 78% of the Australian population fall in this range (based on the NSMH),[42] and this group has one quarter the probability of anxiety or depression compared to the rest of the general population.
- *16-29: medium probability of having anxiety or depression*. 20% of the Australian population fall in this range,[42] and this group has a 25% probability of anxiety or depression.

- *30-50: high probability of having anxiety or depression*. 2% of the Australian population fall in this range,[42] and this group has a 75% probability of anxiety or depression, and a 6% probability of suicide attempt.

In the current study, a much higher proportion of the exposed group had a medium or high probability of having anxiety/depression compared to the comparison groups: 50% versus 29% in Richmond and 32% in Amberley. This translates into a statistically significant 2.5-fold increase in the probability of anxiety/depression; this was consistent in Program 1 and 2 subgroups, and there was a suggestion of a dose-response effect. The proportion of the exposed group in the medium/high category was also much higher than in the Australian population (50% versus 22% respectively).

The GHQ-12 identifies short-term changes in mental health: depression, anxiety, social dysfunction, and somatic symptoms. The score, from 0 to 12, is dichotomised into a low probability category (<4) or high probability category (≥4) of having mental illness. The proportion of the exposed group falling into the high category range was higher than the comparison groups, i.e. 65% versus 49% in Amberley and 47% in Richmond. This represents an odds ratio of approximately 2.2, and is significant in the overall comparison, the Program 1 and 2 subgroups, and shows a clear dose-response relationship. Interestingly, both the exposed and the comparison groups fared much worse than the Australian population. Normative data from the National Survey of Mental Health[42] show that less than 10% of the Australian population scored 4 or greater on the GHQ versus the 48% and 65% recorded in the comparison and DSRS groups respectively.

Neurasthenia is a condition characterised by easy fatiguability and usually associated with non-specific somatic complaints, such as muscular aches, dizziness, tension headaches, sleep problems, inability to relax, and irritability. Neurasthenia was judged present in 9% of the exposed group versus 6% in the Amberley and 3% in the Richmond comparisons. This translates into a statistically significant 2.5-fold increased probability of having neurasthenia in the DSRS group versus the Richmond comparisons only; this is consistent in Program 1 but not significant in Program 2, and there is the suggestion of a dose-response effect. Many of the symptoms included in neurasthenia match the list of common symptoms reported in Chapter 9 (General Health and Well-being), a further indication of the strength and consistency of the data.

The instrument used to assess neurasthenia in this study was identical to that used in the NSMH, and in this case there is a marked increase in the prevalence of the condition

compared to the Australian population. The NSMH reports a prevalence of between 1-2% among middle-aged males, versus 3%, 6% and 9% in Richmond, Amberley, and exposed groups respectively.

15.7 Conclusions

In summary, there is strong evidence of an association between exposure and impaired mental health, particularly anxiety and depression. For these two outcomes, the data for self-reported physician diagnoses, use of medications, and diagnoses on CIDI are remarkably consistent and show a statistically significant doubling of the risk on average. This is generally consistent for both comparison groups, Program 1 and two subgroup analyses, and dose-response effects. Results from the K-10 and GHQ-12 lend further support to these results, indicating a strong association that is consistent across both comparison groups, and both subgroups, and that observes a dose-response effect. Neurasthenia is also significantly increased in the exposed group, although only in comparison to the Richmond group. Data are consistent in subgroup analyses and there is the suggestion of a dose-response effect, and, in addition, many of the symptoms match those self-reported by the groups in the Postal Questionnaire.

15.8 References

1. Donaldson, E. Nature and extent of health complaints, Report to the F111 Deseal/Reseal Board of Inquiry. Volume 2, Part 1, Chapter 13, Annex B. (RAAF, Amberley, 2000).
2. Arlien-Sobørg, P. *Solvent Neurotoxicity*, (CRC Press, Boca Raton, 1992).
3. Axelson, O., Hane, M. & Hogstedt, C. A case-referent study on neuropsychiatric disorders among workers exposed to solvents. *Scandinavian Journal of Work Environment and Health* **2**, 14-20 (1976).
4. Baker, E.L., Smith, T.J. & Landrigan, P.J. The neurotoxicity of industrial solvents: a review of the literature. *American Journal of Industrial Medicine* **8**, 207-217 (1985).
5. Cherry, N., Hutchins, H., Pace, T. & Waldron, H.A. Neurobehavioural effects of repeated occupational exposure to toluene and paint solvents. *British Journal of Industrial Medicine* **42**, 291-300 (1985).

6. Linz, D.H., de Garmo, P.L. & Morton, W.E.e.a. Organic solvent-induced encephalopathy in industrial painters. *Journal of Occupational Medicine* **28**, 119-125 (1986).
7. Morrow, L.A., Ryan, C.M., Hodgson, M.J. & Robin, N. Alterations in cognitive and psychological functioning after organic solvent exposure. *Journal of Occupational Medicine* **32**, 444-50 (1990).
8. Bolla, K.I., Schwartz, B.S., Agnew, J., Ford, P.D. & Bleecker, M.L. Subclinical neuropsychiatric effects of chronic low-level solvent exposure in US paint manufacturers. *Journal of Occupational Medicine* **32**, 671-677 (1990).
9. Escalona, E., Yanes, L., Feo, O. & Maizlish, N. Neurobehavioural evaluation of Venezuelan workers exposed to organic solvent mixtures. *American Journal of Industrial Medicine* **27**, 15-27 (1995).
10. Morrow, L.A. et al. Increased incidence of anxiety and depressive disorders in persons with organic solvent exposure. *Psychosomatic Medicine* **62**, 746-750 (2000).
11. Morrow, L.A., Ryan, C., Goldstein, G. & Hodgson, M.J. A distinct pattern of personality disturbances following exposure to mixtures of organic solvents. *Journal of Occupational Medicine* **31**, 743-746 (1989).
12. White, R.F., Proctor, S.P., Echeverria, D., Schweikert, J. & Feldman, R.G. Neurobehavioural effects of acute and chronic mixed-solvent exposure in the screen printing industry. *American Journal of Industrial Medicine* **28**, 221-231 (1995).
13. Parkinson, D.K. et al. Health effects of long-term solvent exposure among women in blue-collar occupations. *American Journal of Industrial Medicine* **17**, 661-675 (1990).
14. Morrow, L.A., Kamis, H. & Hodgson, M.J. Psychiatric symptomatology in persons with organic solvent exposure. *Journal of Consulting and Clinical Psychology* **61**, 171-4 (1993).
15. Morrow, L.A., Ryan, C.M., Hodgson, M. & Robin, N. Alterations in cognitive and psychological functioning after organic solvent exposure. *Journal of Occupational Medicine* **32**, 444-50 (1990).
16. Stollery, B.T. & Flindt, M.L. Memory sequelae of solvent intoxication. *Scandinavian Journal of Work and Environmental Health* **14**, 45-48 (1988).
17. Laire, G., Viaene, M.K., Veulemans, H., Masschelein, R. & Nemery, B. Nocturnal oxygen desaturation, as assessed by home oximetry, in long-term solvent-exposed workers. *American Journal of Industrial Medicine* **32**, 656-664 (1997).
18. Bowler, R.M. et al. Neuropsychological dysfunction, mood disturbance, and emotional status of munition workers. *Applied Neuropsychology* **8**, 74-90 (2001).

19. Bell, I.R., Baldwin, C.M. & Schottenfeld, R.S. Chapter 31, Psychological Sequelae of Hazardous Materials Exposures. in *Clinical Environmental Health and Toxic Exposures* (ed. Krieger, G.R.) 404-412 (Lippincott Williams & Wilkins, Philadelphia, 2001).
20. Morrow, L.A., Ryan, C.M., Hodgson, M.J. & Robin, N. Risk factors associated with persistence of neuropsychological deficits in persons with organic solvent exposure. *Journal of Nervous and Mental Disease* **179**, 540-5 (1991).
21. WHO. International Classification of Diseases and Related Disorders (ICD-10). (World Health Organisation, Geveva, 1992).
22. Fidler, A.T., Baker, E.L. & Letz, R.E. Neurobehavioural effects of occupational exposure to organic solvents among construction painters. *British Journal of Industrial Medicine* **44**, 292-308 (1987).
23. Kessler, R.C. et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine* **32**, 959-976 (2002).
24. Goldberg, D.P. & Blackwell, B. Psychiatric illness in general practice. A detailed study using a new method of case identification. *British Medical Journal* **1**, 439-443 (1970).
25. WHO. Composite International Diagnostic Interview (CIDI): Core Version 2.1 Interviewer's Manual. (World Health Organization, Geneva, 1997).
26. Saunders, D. & Daly, A. Collaborative Health and Wellbeing Study: Psychological Distress in the Western Australia Population. (Health Department of Western Australia, 2000).
27. NSW Health Department. NSW Health Survey 1997. (Epidemiology and Surveillance Branch, NSW Health Department, Sydney, NSW., 1997).
28. Dal Grande, E., Taylor, A. & Wilson, D. South Australian Health and Wellbeing Study. (Centre for Population Studies in Epidemiology, Department of Human Services, 2002).
29. Andrews, G. & Slade, T. Interpreting scores on the Kessler psychological distress scale (K10). *Australian and New Zealand Journal of Public Health* **25**, 494-497 (2001).
30. Dear, K., Henderson, S. & Korten, A. Well-being in Australia. Findings from the National Survey of Mental Health and Well-being. *Social Psychiatry & Psychiatric Epidemiology* **37**, 503-509 (2002).
31. Clinical Research Unit for Anxiety and Depression, C. K10 Symptom Scale. Vol. 2004 (The University of NSW, http://www.crufad.unsw.edu.au/K10/k10info.htm).
32. Goldberg, D. & Williams, P.A. *Users Guide to General Health Questionnaire.*, (NEER-Nelson, 1988).
33. Goldberg, D. General health questionnaire [GHQ]. in *Measuring Health: A Guide to Rating Scales and Questionnaires.* (eds. McDowell, I. & Newell, C.) 139-150 (Oxford University Press, New York, 1996).

34. Schmitz, N., Kruse, J., Heckrath, C., Alberti, L. & Tress, W. Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social Psychiatry and Psychiatric Epidemiology* **34**, 360-366 (1999).
35. Goodchild, M.E. & Duncan-Jones, P. Chronicity and the General Health Questionnaire. *British Journal of Psychiatry* **146**, 55-61 (1985).
36. Donath, S. The validity of the 12-item General Health Questionnaire in Australia: a comparison between three scoring methods. *Australian and New Zealand Journal of Psychiatry* **35**, 231-235 (2001).
37. Peters, L. & Andrews, G. Procedural validity of the computerized version of the Composite International Diagnostic Interview (CIDI-Auto) in the anxiety disorders. *Psychological Medicine* **25**, 1269-1280 (1995).
38. Rosenman, S.J., Korten, A.E. & Levings, C.T. Computerised diagnosis in acute psychiatry: validity of CIDI-AUTO against routine clinical diagnoses. *Journal of Psychiatric Research* **31**, 581-592 (1997).
39. Rey, A. *L'Examen Clinique en Psychologie*, (Presses Universitaire de France., Paris, 1964).
40. Goldberg, J.O. & MIller, H.R. Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology* **42**, 792-795 (1986).
41. Hiscock, C.K., Branham, J.D. & Hiscock, M. Detection of feigned cognitive impairment: The two-alternative forced-choice method compared with selected conventional tests. *Journal of Psychopathology and Behavioral Assessment* **16**, 95-110 (1994).
42. Andrews, G., Hall, W., Teesson, M. & Henderson, S. The Mental Health of Australians. (Mental Health Branch, Commonwealth Department of Health and Aged Care, Canberra, ACT, 1999).

This page has been left blank intentionally
for formatting purposes

16 Neuropsychological Outcomes

Chapter summary

The General Health and Medical Study included a comprehensive assessment of cognition, which included tests of executive functioning, psychomotor speed, attention, working memory, and new learning. Also, self-reported symptoms of forgetfulness, loss of concentration and difficulty finding the right word were analysed from the Postal Questionnaire. Depression and anxiety (as measured by the CIDI) were included as potential confounders in analyses of all data in this domain. The exposed group scored significantly lower on two tests of executive functioning (the COWAT Letter and Animal tasks) in the overall analysis and had a clear dose-response effect. All three tests of psychomotor speed indicated a statistically significant decrease in performance for the DSRS group for all exposed and Program 1 and 2 subgroup analyses. There were no significant group differences in either of the attention/working memory tests. Only the AVLT test for new learning/memory was sensitive enough to detect differences between the exposed and comparison groups, with the exposed performing worse than Richmond on the Immediate and Delayed Recall and Total Learning tasks. Self-reported memory complaints were significantly increased in the exposed group relative to both comparison groups; this was consistent across Programs 1 and 2 and showed a dose-response effect. Similarly, the exposed group was also more likely to report symptoms of forgetfulness, loss of concentration and difficulty finding the right word, with an increase of two to four times relative to the comparison groups, and with a strong dose-response effect.

Chapter contents

16.1 Introduction

Cognition and memory were a particular focus of the current study, due mainly to the large number of complaints reported in this area to the Board of Inquiry (47% of individuals reported some neurological/psychological symptoms)[1] and to the F-111 Interim Health Care Scheme.

The human nervous system enjoys relative protection from toxic injury; however, compounds that are non-polar and lipid-soluble readily cross the blood-brain barrier. Neurotoxic compounds include arsenic, metals (lead, manganese, mercury and tin), solvents (carbon bisulphide, *n*-hexane, methyl *n*-butyl ketone, perchloroethylene, toluene, trichloroethylene) and insecticides (organophosphate and carbamate).

Two types of neurotoxicities may occur:

1) Acute neurotoxicity typically causes physiological or biochemical changes in the nervous system that involve no structural change. The effect is usually rapidly reversed after withdrawal of the exposure and is particularly associated with organic solvent vapours.

2) Chronic exposures are more often associated with structural changes in the nervous system. These changes are due either to long-standing metabolic derangements that damage nervous tissue, or to hypoxia and ischaemia due to inadequate oxygen;[2] they are typically associated with metals and some high-dose organic solvents.

In 1984, Anger[3] reported that of all the chemicals assessed by the American Conference of Governmental Industrial Hygienists as neurotoxic, at least 25% were organic solvents. Numerous studies have reported reduced neuropsychological performance in subjects exposed to organic solvents including toluene[4-6] and solvent mixtures.[7-13] Organic compounds in jet fuel have also been associated with neuropsychological effects[14,15] and more recently with poor performance on attention tasks and information processing speed.[16]

The neuropsychological effects of neurotoxins in adults are usually assessed by impairments in one or more of the following seven functional areas: attention, executive functioning, fluency (verbal or visual), motor abilities, visuospatial abilities, learning and short-term memory, and mood and adjustment.[3] To assess these domains in a standardised manner, “test batteries” have been established, e.g. the World Health Organisation (WHO) and the

US National Institute of Occupational Safety and Health (NIOSH). From these and other batteries, the current study selected a series of tests that would assess cognition and memory both globally as well as in the specific domains of executive functioning, psychomotor speed, attention/working memory, new learning/memory, visuospatial ability, subjective memory complaints, and self-reported symptoms.

16.2 Measures

16.2.1 Executive functioning

Executive functioning is the ability to organise thoughts and work, to create and successfully execute plans, and to manage administrative functions that are part of day-to-day living while adjusting one's actions along the way as needed. Executive functioning can be assessed both informally and formally by using clinician observation of patient management of different real-world situations alongside formal structured assessments which provide standardised scores. The executive functioning domain was assessed through the use of three tests during the health examination:

(1) The Wechsler Adult Intelligence Scale (WAIS III) similarities test forms part of the "Verbal Scale" aspect of the overall WAIS III. It required the participant to describe how two given objects were alike. It measured concrete, functional and abstract concept formation. Each participant was told two words that represented common objects or concepts. They then had to state how the two objects or concepts were alike. Each response was scored 0, 1 or 2 depending upon the quality of the answers. The participant was asked to discontinue after four consecutive scores of zero.

(2) The WAIS III Controlled Oral Word Association verbal fluency test (COWAT) was used to assess verbal fluency and the ease with which a person could think of words that began with a specific letter. The HSA psychologist instructed the participant to verbalise as many words as possible that began with a particular letter as quickly as they could, once the particular letter had been stated. An example using the letter "T" was given, with the participant asked to respond with four or five words of their own. Proper names were not permitted, such as Texas, Townsville or Toyota, nor could the participant use the same word with a different ending, such as "teeth" followed by "teething". The first test letter given to the participant was

“F”, followed by “A” and “S”, and finally they were asked for as many names of animals as they could think of (that began with any letter). Each letter/animal test lasted for 60 seconds. Each correct word and animal scored 1 point each, with each of the 4 sub-scores being totalled overall (F + A + S + animals). SHOAMP analysed the sum of the three letter tests, then the animal test separately.

(3) The Trail Making Test (TMT) is a well-established test sensitive to impairment in multiple cognitive domains. It consists of Parts A and B, with Part B reported as being more sensitive than Part A to impairments in cognitive flexibility. Part B is more difficult to complete due to its extra length and its having more than one item in the path of the trail which creates visual interference. SHOAMP participants were required to draw a line to connect, in order, a series of numbers and letters (1-A-2-B-3-C etc) as quickly as they could. The first test of Part B was relatively short, with the second series of numbers and letters being longer. Each test was timed, with the total time and the number of errors per test being recorded. The majority of participants were expected to complete Part B within three minutes, with > 273 seconds indicating a deficiency.

16.2.2 Psychomotor speed

Psychomotor speed is the ability to rapidly and fluently perform body motor movements. Three tasks were administered as tests of psychomotor speed: WAIS III Digit Symbol Coding, Trail Making Test Part A, and the Purdue Pegboard.

(1) The WAIS III can be used for the assessment of learning disabilities.[17] The Digit Symbol Coding (DSC) performance sub-test of the WAIS III requires subjects to demonstrate visuo-motor speed and scanning accuracy. For Digit Symbol Coding, SHOAMP participants had to copy symbols that were paired with numbers. Using a key, the participant drew each symbol under its corresponding number. The participant’s score was determined by the number of symbols correctly drawn within the 120-second time limit, with a higher score meaning more symbols had been copied correctly. A series of sample items were provided to each participant to ensure they understood the task. Participants were encouraged to work through the task as quickly as they could, with spontaneous corrections allowed (no marks were deducted). Each item had to be completed in order, with no skipping. Participants received one point for each correctly drawn symbol completed within the 120-second time limit (excluding sample items). Credit was not given for items completed out of sequence. A response was deemed correct if it was clearly identifiable as the keyed symbol, even if it was

drawn imperfectly or included a spontaneous correction of an incorrect symbol. The maximum score for a participant was 133 points.

(2) Part A of the TMT is administered before Part B, is not as long and contains only numbers without letters. TMT Part A involved two tests of different lengths, where each participant was required to connect numbers in order from 1 onwards to “end”, while being timed. Total time and number of errors per test were recorded. The majority of participants were expected to complete Part A within 90 seconds, with > 78 seconds indicating a deficiency.

(3) The Purdue Pegboard[18] is a standardised test used for gross movements of hands, fingers and arms, and for assessing fine motor skills, in particular finger dexterity and hand-eye coordination. The test consists of picking up small steel pegs from a well in a pegboard and placing them sequentially in 10 holes as quickly as possible. For this test each participant received one practice and one trial only for each of right hand, left hand, both hands, and assembly conditions (the four test procedures). Participants were required to place as many pins as possible in the holes on the pegboard; firstly with the preferred hand, then with the other hand, and finally with both hands, within a 30-second time limit for each condition. After each procedure, the pins were removed and the board was prepared for the next test. The fourth procedure (the assembly) had a 60-second time limit. For each participant, five scores were generated, based on the number of pins inserted into the board for each of the following task types:

- pins left hand (30 secs)
- pins right hand (30 secs)
- pairs of pins, both hands (30 seconds)
- the sum of left hand, right hand, and both hands
- parts assembled (60 secs).

For the “both hand” task, the score reflected the pairs of pins inserted, not the total number. The total score was obtained from summing the test scores of the previous three test batteries. For the “assembly condtions” score, there were four parts in each assembly, so the total number of complete assemblies were multiplied by four and then parts from incomplete assemblies were added.

16.2.3 Attention/working memory

This domain is the ability to integrate and manipulate new data. Two tests were included for SHOAMP:

(1) The WAIS III Digit Span test was used as a test of immediate auditory recall and freedom from distraction. Lower scores indicate an attention deficit or anxiety. Digit Span is composed of two tasks administered independently of each other: Digits Forward task and Digits Backward task. On both tasks, the attending HSA psychologist read a series of number sequences to the participant. For each Digits Forward task, the participant was required to repeat a number sequence in the same order as was presented to them (e.g. repeat the numbers 1, 7 or repeat the numbers 6, 4, 3 and 9 in order). There were eight Forward tasks to work through, each comprising two parts (which were scored as “0” for a fail or “1” for a pass, giving each of the eight tasks a possible score of “2” if both parts were correct). A total score of 16 is possible for Digit Span Forwards if each number sequence was correctly answered by the participant.

(2) For Digits Backward, the participant was required to repeat a number sequence in the reverse order (e.g. numbers read out as 1, 7 would be answered as 7, 1). There were seven Backwards tasks to work through, each comprising two parts (which were scored as “0” for a fail or “1” for a pass, giving each of the seven tasks a possible score of “2” if both parts were correct). A total score of 14 was possible for Digit Span Backwards if each number sequence was correctly answered by the participant.

16.2.4 New learning/memory

This domain is the ability to absorb, store and recall new data after a delay. New learning and memory was assessed by six tests administered during the health examination in addition to a self-reported subjective memory complaint survey included in the Postal Questionnaire.

(1) From the WAIS III, two sub-tests were included as tests of non-verbal memory: (a) Digit Symbol Incidental Learning Pairing, and (b) Digit Symbol Incidental Learning Free Recall. The Digit Symbol Coding task (previously described as part of executive functioning) was administered first, then immediately after that test was completed each participant was exposed to both remaining parts of the Incidental Learning task.

(2) Incidental Learning (Pairing) consisted of two rows of nine items, which were numbers without symbols. The participant was required to fill in all the symbols that matched the numbers. One point was recorded for each correct Pairing response, up to a maximum score of 18. For Incidental Learning (Free Recall) the participant was required to remember as many of the symbols as possible and record them in any order (maximum score was 18).

(3) The WMS Visual Reproduction memory assessment task involved the attending HSA psychologist showing the participant a series of four drawings on three stimulus cards and asking the participant to redraw each design one at a time after seeing it briefly for 10 seconds. A modified scoring procedure was used, as suggested by Ryan et al.,[19] which included provision for an additional point to be given for each design, based on accuracy of spatial relationships. Thus a total of 17 points was possible for both Immediate and Delayed Recall (see below).

(4) A Delayed Recall test was also administered to participants, where they were asked to recall their drawn designs after they had completed other tests. The participant was asked to recall the drawings previously shown to them on cards by the psychologist. Some clues were provided if the person could not recall any item at all. For each of the four designs being recalled, Ryan's scoring system[19] (below) was applied, with the design elements having to be present to receive accuracy points.

Revised scoring criteria

Design A: 4 points possible

* 1 point if 2 lines crossed
* 1 point if 4 flags drawn
  1 point if flags facing correctly
  1 point for accuracy. Lines are nearly equal (within 1 cm), are nearly bisected, are nearly at right angles *and* flags are nearly square.

Design B: 5 points possible

* 1 point if large quadrilateral drawn with two large diameters
* 1 point if 4 small quadrilaterals drawn (independent of 2 large diameters)
  1 point for 2 small diameters *in each small quadrilateral*
  1 point for 16 dots, each alone in a small quadrilateral

1 point for accuracy. Largest width around is less than or equal to width of largest small quadrilateral, and the two large diameters are equal in length (within 1 cm).

Design C1: 4 points possible

1 point if left loop faces centre

1 point if centre open at bottom

1 point if right loop faces centre

1 point for accuracy. Height of centre is less than or equal to top edge *and* is greater than or equal to the bottom edge of the side loops. The side loops are nearly square (within ½ cm) and nearly symmetrical.

Design C2: 4 points possible

1 point for large quadrilateral with small quadrilateral inside

1 point if all vertices are connected

1 point if centre quadrilateral is shifted to right

1 point for accuracy. Both quadrilaterals must be clearly rectangular with parallel *and* symmetrical sides. Width of centre quadrilateral must be greater than or equal to width of left space.

(5 & 6) The Auditory Verbal Learning Test (AVLT) is a word-list learning task in which fifteen words (List A) were presented five times, with recall tested after each presentation; these were Trials I to V. A second list was then presented once, with immediate recall of that list tested (List B); this was Trial VI. To assess the impact of having to learn a second list (an interference task) on a person's memory for the first list, participant recall for the first list of words was tested next, but without cues (Trial VII). Approximately 20 minutes later, recall for the words in the first list was again tested unexpectedly (Trial VIII). Finally, 50 words were read out with the participant asked to identify the words that had been read out earlier (Recognition Memory). Fifteen words were from List A, 15 words from List B, and 20 words were from neither list. Words could be matched correctly to the list of origin, they could be assigned to the wrong list (misplaced), or non-list words could be incorrectly recognised as belonging to List A or List B (intrusions).

The AVLT responses provided by each participant were entered into the computer-scored Geffen Program (AVLT Scoring Package AVLT version 3.0). Once entered, responses were then scored by the program. The Geffen output consisted of three files per participant, with file extensions of “avl”, “anl” and “txt”. Of these, the “anl” file was formatted for use by statistical programs. Once all the Auditory Verbal Learning Tests had been entered into the

Geffen program and scored, the .anl files were then all grouped together. A number of different measures were derived using the different aspects of memory function. The score for each trial was the number of words correctly recalled. In addition to scores on AVLT Trials I to V, which may be used to plot a learning curve, the AVLT yields scores for the total number of words recalled following interference (post-distraction Trial or Trial VI), the number of words recalled after the 20-minute delay, and the total number of words recognised from each list. Other scores, including a total score (the sum of Trials I to V), the number of repetitions and extra-list intrusions, and the amount of loss from Trial V to the post-distraction Recall Trial (VI) were calculated.

16.2.5 Visuospatial ability

Spatial problem-solving and manipulative abilities, and part-to-whole organisation, were assessed with the WAIS III Block Design test. The test material included nine coloured blocks, each with two white sides, two red sides and two half-red and half-white sides, and a booklet illustrating different colour designs that can be formed. For this test the participant was asked to replicate models or pictures of two-colour designs with blocks. The designs progressed in difficulty from simple two-block designs to more complex, nine-block designs. For each design test, time restrictions ranged from 60 to 120 seconds. A maximum raw score of 68 could be achieved, with a higher score indicating better performance. For Block Design test items 1 to 6, a person scored up to two points for each correct design or zero points for each incorrect design (maximum score 12 points). For test items 7 to 14, the score was based upon the amount of time taken to replicate each design, with scores of 4, 5, 6 or 7 possible for each design test depending upon the period of time that applied (which differed between tests).

16.2.6 Subjective memory complaints

This is the subject's self-reported assessment of any memory problems. Although some studies indicate that subjective reports of cognitive difficulties, such as memory problems, do not always correlate with objective data,[20-22] other studies indicate that, in general, people's assessment of their own memory abilities corresponds to their actual performance on cognitive measures.[23] The six-item MAC-Q was designed to quantify subjective memory complaints with scores ranging from 7 to 35. Its authors established a cut-off point of 25 or above out of 35, with a higher score representing poorer performance.[24] The questionnaire contains five items (each scored from 1 to 5 points) which address daily activities, and one

question (the final item scored as either 2, 4, 6, 8 or 10 points) which addresses overall memory functioning by comparing the present to when the person was 18 to 20 years old. Participants were invited to choose one of the five options per item, ranging from "*much better now"* to "*much worse now".* In previous studies, the MAC-Q has demonstrated satisfactory internal consistency, test/re-test reliability and concurrent validity,[24] and provides SHOAMP participants with the opportunity to express their own memory problems.

16.2.7 Global measure of cognitive function

The MMSE is a widely-used standardised cognitive screening test. It was first described by Folstein in 1975 as a "practical method for grading the cognitive state".[25] MMSE has been standardised for different languages and cultures[26-29], and the reliability and construct validity have been judged to be satisfactory[30] with O'Connor et al.[26] reporting a sensitivity of 86% (those judged to have organic mental disorders) and a specificity of 92% (those judged to be cognitively intact) when using the cut off score of ≤23 out of 30. The advantages of the MMSE include its brevity (5-10 minutes to administer), and the fact that it is a global assessment of many domains including orientation to time and place (10 points), registration of three words (3 points), attention and calculation (5 points), recall of three words (3 points), language (8 points), and visual construction (1 point).

Folstein's Mini Mental State Examination[25] (MMSE) was included as part of an overall assessment of mental state of participants in the General Health and Medical Study. The MMSE is a brief, quantitative measure of cognitive status in adults. It can be used to screen for cognitive impairment, to estimate the severity of cognitive impairment at a given point in time, to follow the course of cognitive changes in an individual over time, and to document an individual's response to treatment. The MMSE has also been used as a research tool to screen for cognitive disorders in epidemiological studies and to follow cognitive changes in clinical trials.

As a global measure of cognitive function, the MMSE was administered to participants by the attending HSA doctor. Scores on the MMSE range from 0-30, with a higher score indicating better overall performance. A score below 24 out of 30 indicates probable cognitive impairment. A score below 21 out of 30 indicates definite cognitive impairment. Each participant was told that they were doing a "memory test" and that there were no time limits. Each person was asked to answer two questions assessing orientation, one item assessing registration (i.e. immediate recall of words), two items assessing concentration (of which the

best score for one only is kept), one item assessing short-term memory (to recall items from registration test), five items assessing language and praxis, and one item assessing visuospatial abilities.

16.2.8 Self-reported symptoms

Three items referring to general cognition were summarised from the Postal Questionnaire. Participants were asked to respond either yes or no to having experienced “forgetfulness” (item 2.30), “loss of concentration” (item 2.36) or “difficulty finding the right word” (item 2.37) in the past month. Each of these items were reported individually, without being combined.

16.3 Potential confounders

As with previous analyses, basic potential confounders across all domains were age (categorised into five-year intervals), posting and rank. Additionally, other potential confounders were alcohol intake (categorised), smoking behaviour, educational background, and civilian exposure to organic solvents before enrolment in the forces. Where HSA centre was found to be a significant predictor, centre was included as a potential confounder.

Depressive symptoms have been reported to be associated with a greater likelihood of reporting memory impairment.[23] It has also been noted that people minimise actual memory decline who are functioning well in activities of daily living and are not suffering depressive symptoms,[23] and that subjective memory problems usually improve if depression is alleviated.[31] Tobiansky et al.[32] reported that within their sample of elderly residents in an electoral hospital ward, 25% of subjects reported subjective memory impairment, with impairment more likely to be reported by those suffering from dementia or depression. For this set of analyses, depression and anxiety were also included as potential confounders (as measured by CIDI scores; see Chapter 15). In regression analyses, depression has been found to be a significant predictor of cognitive impairment,[33,34] hence depression and anxiety were included as covariates in the analysis.

16.4 Analyses

All analyses followed the outline described in Chapter 6 (Analysis). Points of note include:

- Scores for Trail Making Tests Part A and B required a log transform to satisfy analysis assumptions.
- For the Purdue Pegboard, it was decided that the most appropriate outcome variable was the sum of scores for the left, right, and both hands tasks. Examination of the "number of parts assembled" variable indicated that there may have been some inconsistency in recording, with only some examiners conducting the appropriate multiplication (by four) to arrive at the final score; thus this variable was excluded from analysis.
- Where distributions of the outcome measures were not suitable for linear model analysis, the outcome was dichotomised around the 10^{th} percentile within each age group of the Richmond comparison group, and comparisons were performed using logistic regression.
- The scores on the Rey 15-item test were used to screen out those whose test scores might not be reliable.

In addition to health examination data, several self-reported Postal Questionnaire items were examined: "forgetfulness", "loss of concentration" and "difficulty finding the right word", in the past month. For these items, primary analysis involved multiple logistic regression using exposure group and all potential confounders combined as explanatory variables, and secondary analyses were performed using multiple logistic regression separately for Program 1, Program 2, and for a dose-response.

16.5 Results

16.5.1 Rey 15-item test

The Rey 15-item test was completed by 1532 participants, which represents 99.6% of those participating in the General Health and Medical Study. Of those not completing the test, two refused and four did not undertake any psychological testing. Of those completing the test, 75% completed all 15 letters, shapes and symbols correctly. There were 17 participants

whose Rey test score fell below the cut-off point of less than 8, with 1.2% from the Amberley comparison group, 0.6% from the Richmond comparison group, and 1.5% from the exposed group.

16.5.2 Executive functioning

WAIS III Similarities

Of the 1538 eligible participants in the health study, one was excluded due to stroke, 12 did not undertake the similarities test, and a further 17 were excluded due to their poor performance on the Rey test. This left a total of 1508 for analysis. Similarities was analysed as total raw score, and the range in this sample was from 8 to 33. The distribution was acceptably normal and did not require transformation. Average similarities raw scores varied significantly ($p<0.0001$) across HSA centres but there was no evidence of heterogeneity ($p=0.39$); in other words there was no evidence to suggest that group differences varied across centres. HSA centre was included in the analysis as a potential confounder.

The mean total raw score for the exposed group fell between those of the two comparison groups at 23.8 (see Table 16.1).

Table 16.1 : WAIS Similarities – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	22.90	24.49	23.79
Standard Deviation	5.33	4.28	4.52
50th Percentile	23.00	25.00	24.00
Lower Quartile	20.00	22.00	21.00
Upper Quartile	27.00	28.00	27.00

Although group appeared significant in the multivariable linear model (Table 16.2) for all exposed ($p=0.0034$), Program 1 ($p=0.02$) and Program 2 ($p=0.008$), most of this was driven by Amberley versus Richmond comparisons. Although the exposed group appeared significant versus Richmond in the overall analysis for the full and reduced models, this was lost in the Program 1 and Program 2 sub-analyses and with the robust standard variances estimates. There was no dose-response effect ($p=0.23$).

Table 16.2 : WAIS Similarities – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R square	DF	p-value
All Exposed	Amberley	-0.45	-1.11	0.21	0.17	2	0.0034
	Richmond	0.66	0.08	1.24			
Program 1	Amberley	-0.57	-1.40	0.26	0.18	2	0.0203
	Richmond	0.42	-0.32	1.17			
Program 2	Amberley	-0.55	-1.41	0.31	0.19	2	0.0084
	Richmond	0.53	-0.26	1.32			
Dose	Mild exposure	0.45	-0.33	1.24	0.17	3	0.2271
	Moderate exposure	-0.42	-1.19	0.34			
	Prolonged exposure	-0.40	-1.16	0.35			

WAIS III Controlled Oral Word Association Test (COWAT)

In total, 11 participants did not undertake the COWAT Letter test in whole or in part. As before, a further 17 were excluded due to poor performance on the Rey test, and one individual was excluded because of impairment due to a previous stroke, leaving 1509 available for analysis. The COWAT Letter task was analysed as the sum of the raw scores for the three letters F, A and S, and these total scores ranged from 11 to 79. Whilst there was significant variation in scores across HSA centres (p=0.0004) there was no heterogeneity observed (p=0.73).

Unadjusted combined raw scores were similar for Amberley and Richmond comparisons at 39 and 40 respectively, while the score for the exposed group was two to three points lower at 37 (see Table 16.3).

Table 16.3 : COWAT Letter Task – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	39.16	39.71	37.12
Standard Deviation	10.91	10.90	10.55
50th Percentile	39.00	40.00	36.00
Lower Quartile	32.00	32.00	30.00
Upper Quartile	46.00	47.00	43.00

Following adjustment for all potential confounders (see Table 16.4), there was a significant effect for group when comparing scores between Richmond and all exposed and between Amberley and all exposed (p=0.02), but not when the two comparison groups were compared with exposed from Program 1 or with exposed from Program 2.

Table 16.4 :COWAT Letter Task – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

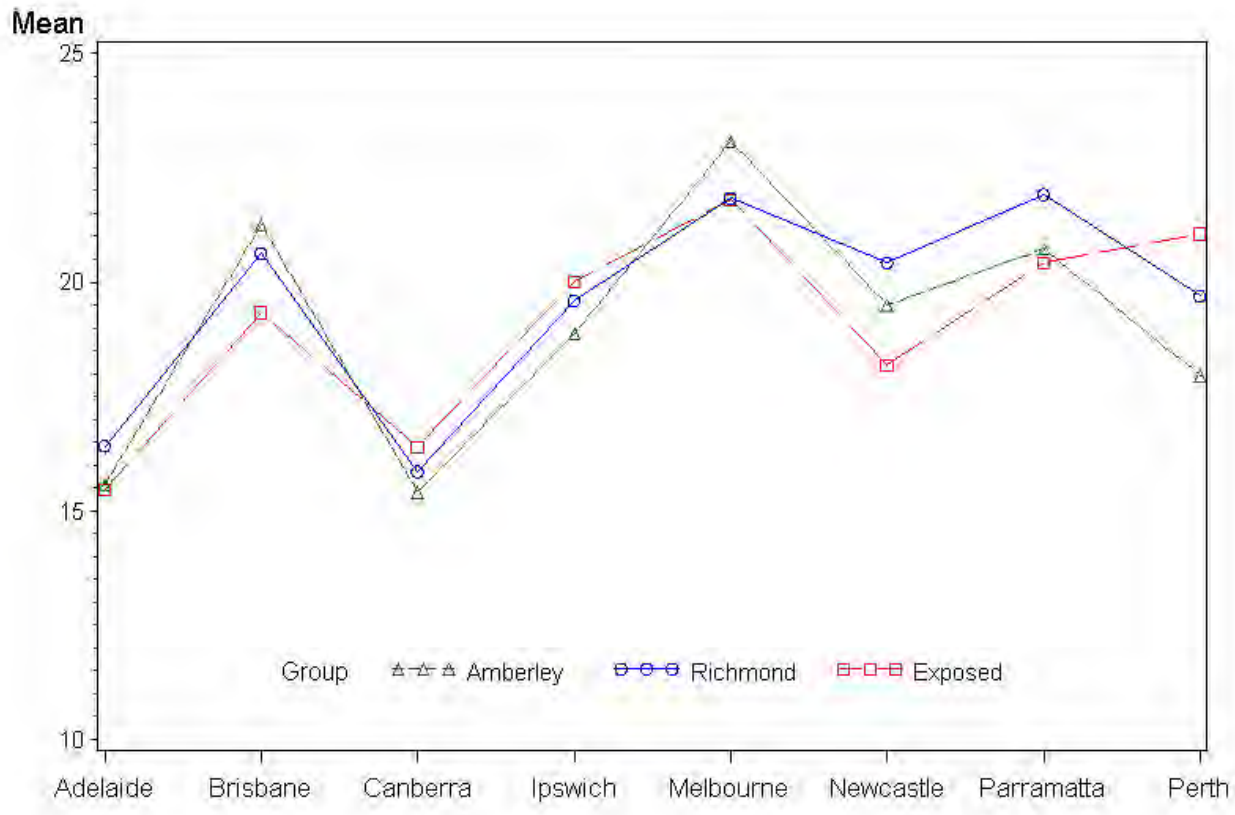
Analysis	Parameter	Estimate	Lower CL	Upper CL	R square	DF	p-value
All Exposed	Amberley	1.72	0.14	3.29	0.10	2	0.0203
	Richmond	1.79	0.41	3.17			
Program 1	Amberley	1.22	-0.71	3.15	0.11	2	0.3021
	Richmond	1.32	-0.42	3.06			
Program 2	Amberley	0.81	-1.24	2.87	0.11	2	0.5503
	Richmond	1.05	-0.84	2.95			
Dose	Mild exposure	-1.35	-3.22	0.51	0.10	3	0.0522
	Moderate exposure	-1.79	-3.61	0.04			
	Prolonged exposure	-2.12	-3.92	-0.31			

When considering the all exposed comparison, Richmond was on average 1.8 points higher than the exposed (95%CI: 0.41, 3.17), and Amberley was on average 1.7 points higher than the exposed (95% CI: 0.14, 3.29). The results remained significant in the reduced model and

with the robust standard error estimates. Dose was just significant in the model (p=0.05) and there was stepwise increase in the estimates. Those who were exposed for less than nine months were on average 1.4 points lower than the unexposed (95% CI: -3.22, 0.51); those who were exposed for nine to 29 months were 1.8 points lower than the unexposed (95% CI: -3.61, 0.04); and those who were exposed for 30 months or more were on average 2.1 points lower than the unexposed (95% CI: -3.92, -0.31).

Nine participants did not complete the COWAT Animal task, and three participants refused to undertake the test. As before, 17 were removed from analysis due to poor results on the Rey test, and one was excluded because of a previous stroke. Consequently there were 1508 results available for analysis. Significant heterogeneity was observed (p=0.016, see Figure 16.1).

Figure 16.1 : COWAT Animal Task heterogeneity



In view of the fact that this is a well-standardised test, and that tests administered at the same time and in the same battery were *not* heterogeneous, this was interpreted as a statistical artifact, and analysis was continued. Unadjusted combined raw scores were similar

for all three groups, with a mean score of 20 for Amberley and Richmond comparison groups and 19 for the exposed group (see Table 16.5).

Table 16.5 : COWAT Animal Task – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	20.23	20.38	19.47
Standard Deviation	5.15	4.77	4.68
50th Percentile	20.00	20.00	20.00
Lower Quartile	17.00	17.00	16.00
Upper Quartile	24.00	24.00	23.00

Following adjustment for all potential confounders (see Table 16.6), there was a significant effect for group when comparing scores between Richmond and all exposed and between Amberley and all exposed (p=0.0003).

Table 16.6 : COWAT Animal Task – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R square	DF	p-value
All Exposed	Amberley	1.10	0.42	1.78	0.20	2	0.0003
	Richmond	1.09	0.49	1.68			
Program 1	Amberley	0.78	-0.05	1.62	0.22	2	0.1121
	Richmond	0.74	-0.01	1.49			
Program 2	Amberley	0.87	-0.01	1.74	0.22	2	0.0838
	Richmond	0.86	0.05	1.67			
Dose	Mild exposure	-0.99	-1.79	-0.19	0.20	3	0.0015
	Moderate exposure	-1.08	-1.87	-0.30			
	Prolonged exposure	-1.21	-1.98	-0.43			

These differences were not significant in Program 1 or Program 2 (p=0.11 and p=0.08, respectively) although the point estimates were similar. When considering the all exposed comparison, Richmond and Amberley were on average both one point higher than exposed (95% CI: 0.5, 1.7 and 95% CI: 0.4, 1.8). This remained consistent in the reduced model, although the Amberley comparisons lost significance with the robust standard error estimates. Dose was a significant effect in the model (p=0.002), with a stepwise increase in the estimates. Those who were exposed for less than nine months were on average one point lower than the unexposed (95% CI: -1.79, -0.19); those who were exposed for nine to 29 months were 1.1 points lower than the unexposed (95% CI: -1.9, -0.30); and those who were exposed for 30 months or more were on average 1.2 points lower than the unexposed (95% CI: -2.0, -0.43).

Trail Making Test Part B

Seventeen health study participants did not undertake the Trail Making Test (TMT) Part B, and three participants refused to undertake the test at all. As before, 17 participants were excluded because they did not pass the Rey test and one because of a previous stroke. One other value was discarded as being out of range (coding error). There were 1499 results available for analysis. Time to complete the test was recorded in seconds, was not normally distributed, and exploratory analysis revealed this was reflected in non-normal distribution of error terms. Consequently a log transformation was applied. A test of group–HSA centre interaction revealed no heterogeneity (p=0.68). Mean log time scores varied significantly across HSA centres; therefore centre was included as a potential confounder in the following analyses. Mean time to complete the test is presented in Table 16.7, and log of mean time is presented in Table 16.8.

Table 16.7 : Trail Making Test Part B – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	65.69	61.97	66.49
Standard Deviation	25.62	23.57	24.98
50th Percentile	62.00	57.00	61.50
Lower Quartile	48.00	48.00	49.00
Upper Quartile	78.00	70.00	77.00

Table 16.8 : Log transformation of Trail Making Test Part B completion time – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	4.12	4.07	4.14
Standard Deviation	0.35	0.33	0.34
50th Percentile	4.13	4.04	4.12
Lower Quartile	3.87	3.87	3.89
Upper Quartile	4.36	4.25	4.34

The unadjusted mean raw scores were much lower for Richmond at 62 seconds, compared with Amberley and the exposed group at 66 seconds. After adjustment for all potential confounders (see Table 16.9), there was a notable group effect driven by the comparison between Richmond and exposed groups in the overall analysis (p=0.0005), Program 1 (p=0.0009) and Program 2 (P=0.0009).

Table 16.9: Trail Making Test Part B – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R square	DF	p-value
All Exposed	Amberley	0.01	-0.04	0.05	0.19	2	0.0005
	Richmond	-0.07	-0.11	-0.03			
Program 1	Amberley	-0.00	-0.06	0.05	0.20	2	0.0009
	Richmond	-0.08	-0.13	-0.03			
Program 2	Amberley	-0.00	-0.06	0.06	0.17	2	0.0009
	Richmond	-0.08	-0.14	-0.03			
Dose	Mild exposure	0.02	-0.04	0.07	0.17	3	0.0771
	Moderate exposure	0.05	-0.00	0.11			
	Prolonged exposure	0.06	0.01	0.11			

Log time to complete the test was significantly lower on average for Richmond (-0.07, 95% CI: -0.11, -0.03) when compared with all exposed, and this remained significant in the reduced model and in robust standard error estimates. No dose-response effect was seen (p=0.08).

16.5.3 Psychomotor speed

WAIS III Digit Symbol Coding

Of the 1538 eligible participants in the Health Study, one was excluded due to a previous stroke, eight did not undertake Digit Symbol Coding, and a further 17 were excluded due to their poor performance on the Rey test. This left a total of 1512 for analysis. Digit Symbol Coding was analysed as a raw score and the range in this sample was from 21 to 173. The distribution was acceptably normal and did not require log transformation. Average Digit Symbol Coding raw scores did not vary significantly (p=0.05) across HSA centres and there was no evidence of heterogeneity (p=0.19); in other words there was no evidence to suggest that group differences varied across centres.

Unadjusted raw scores varied on average between the three groups, with the exposed group scoring about 4 points less (group mean) than the Richmond comparison group and around 1.5 points less than the Amberley comparison group (see Table 16.10).

Table 16.10 : Digit Symbol Coding – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	70.41	72.89	68.97
Standard Deviation	13.97	13.15	14.92
50th Percentile	71.00	73.00	69.50
Lower Quartile	60.00	64.00	59.00
Upper Quartile	81.00	82.00	78.00

After adjustment for all potential confounders in a multivariable linear model (see Table 16.11), a consistently strong significant group effect was observed when comparing Richmond and Amberley to all exposed (p=0.0003), to those in Program 1 (p=0.0009), and to those in Program 2 (p=0.0012). The difference was significant and consistent with respect to Richmond versus the exposed, with a three to four point gain on average for the Richmond group compared to the exposed for all exposed (3.4, 95% CI: 1.7, 5.0), for Program 1 (3.67, 95% CI: 1.6, 5.7) and for Program 2 (3.7, 95% CI: 1.4, 5.9).

Table 16.11 : Digit Symbol Coding – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	1.02	-0.88	2.91	0.25	2	0.0003
	Richmond	3.38	1.72	5.04			
Program 1	Amberley	1.43	-0.86	3.72	0.27	2	0.0009
	Richmond	3.74	1.68	5.81			
Program 2	Amberley	1.11	-1.34	3.57	0.22	2	0.0012
	Richmond	3.78	1.51	6.05			
Dose	Mild exposure	-2.10	-4.35	0.14	0.24	3	0.0025
	Moderate exposure	-3.33	-5.52	-1.14			
	Prolonged exposure	-3.16	-5.33	-0.99			

However, this effect was not apparent when contrasting the Amberley comparison group with the exposed where the 95% confidence intervals spanned zero for all exposed, Program 1 and Program 2. Additionally, a moderate dose effect was observed (p=0.0025) with apparent trend from a 2.10 point drop (95% CI: -4.35, 0.14) for those with up to nine months' exposure compared with those having no exposure, a 3.3 point drop (95% CI: -5.52, -1.14) for those with nine to 29 months exposure compared with no exposure, and a 3.16 point drop (95% CI: -5.33, -0.99) for those with 30 months or more exposure compared with those having no exposure. Results with the reduced model and with the robust standard error estimates were similar.

Trail Making Test Part A

There were 1497 participants who had complete data for the Trail Making Test Part A. Differences between the three groups were in the fractions of seconds (Table 16.12). The distribution of Trail Making completion times was skewed to the right, so means, medians and quartiles are shown for both the completion time and the logarithm of time (the variable used for analysis) (see Table 16.13). The tables indicate that the three groups performed similarly in both tests. There was no evidence of heterogeneity between HSA centres (p=0.89).

Table 16.12 : Trail Making Test Part A completion time – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean (sec)	27.42	26.21	27.61
Standard Deviation	8.72	7.58	8.59
50th Percentile	26.00	25.00	26.00
Lower Quartile	21.00	21.00	22.00
Upper Quartile	32.00	30.00	31.00

Table 16.13 : Log transformation of Trail Making Test Part A completion time – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean (sec)	3.26	3.23	3.27
Standard Deviation	0.31	0.27	0.29
50th Percentile	3.26	3.22	3.26
Lower Quartile	3.04	3.04	3.09
Upper Quartile	3.47	3.40	3.43

Multiple linear regression analyses for Trail Making Test Part A are summarised in Table 16.14. The effect for group was significant in the overall analysis (p=0.005), Program 1 (p=0.018) and Program 2 (p=0.018). Most of this significance was driven by the Richmond versus exposed comparison. Richmond scored 0.05 log units (or ~1 second) faster than the overall exposed group (95% CI: -0.09, -0.02), 0.04 log units lower than Program 1 exposed (95% CI: -0.08, 0.01) and 0.05 log units lower than Program 2 exposed (95% CI: -.010, 0.0). This was similar and significant in the reduced model and with the robust standard error estimates. There was no dose-response effect (p=0.36).

Table 16.14 : Log of completion time for Trail Making Test Part A – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R square	DF	p-value
All Exposed	Amberley	0.00	-0.04	0.05	0.17	2	0.0047
	Richmond	-0.05	-0.09	-0.02			
Program 1	Amberley	0.02	-0.03	0.07	0.17	2	0.0182
	Richmond	-0.04	-0.08	0.01			
Program 2	Amberley	0.01	-0.04	0.06	0.15	2	0.0184
	Richmond	-0.05	-0.10	0.00			
Dose	Mild exposure	0.04	-0.01	0.09	0.16	3	0.3633
	Moderate exposure	0.02	-0.03	0.07			
	Prolonged exposure	0.02	-0.03	0.07			

Purdue Pegboard

1505 participants had complete information for the Purdue Pegboard. Mean scores were identical for the exposed and Amberley groups at 38.4 but slightly higher for the Richmond group at 38.9 (Table 16.15). There was no heterogeneity between centres (p=0.97).

Table 16.15 : Sum of Purdue Pegboard pins inserted with right, left and both hands – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	38.39	38.88	38.45
Standard Deviation	5.21	4.78	5.61
50th Percentile	39.00	39.00	38.00
Lower Quartile	35.00	36.00	35.00
Upper Quartile	41.00	42.00	42.00

Correcting for all potential confounders in the multivariate linear regression indicated a significant (or borderline significant) group effect for all exposed (p=0.049), Program 1 (p=0.016) and Program 2 (p=0.054) (see Table 16.16). Most of this significance was due to

the exposed versus Richmond comparison, with Richmond scoring 0.69 pins more than all exposed (95% CI: 0.04, 1.33), 1.06 pins more than the Program 1 exposed (95% CI: 0.25, 1.86) and 0.75 pins more than the Program 2 exposed, although the last was not significant (95% CI: -0.11, 1.61). These results remained significant in the reduced model and with the robust standard error estimates. There was no clear dose-response effect.

Table 16.16 : Sum of Purdue Pegboard pins – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R square	DF	p-value
All Exposed	Amberley	-0.05	-0.72	0.63	0.18	2	0.0491
	Richmond	0.69	0.04	1.33			
Program 1	Amberley	0.33	-0.50	1.16	0.19	2	0.0161
	Richmond	1.06	0.25	1.86			
Program 2	Amberley	-0.01	-0.88	0.87	0.17	2	0.0542
	Richmond	0.75	-0.11	1.61			
Dose	Mild exposure	-0.20	-1.07	0.67	0.17	3	0.5695
	Moderate exposure	-0.56	-1.41	0.29			
	Prolonged exposure	-0.37	-1.21	0.47			

16.5.4 Attention/working memory

Eight participants did not undertake the Digit Span Forwards or Backwards test. A further 17 participants were excluded from analysis as they did not pass the Rey test, and one participant was excluded because of a previous stroke, leaving 1512 results available for analysis. There was no heterogeneity of group between centres for either the forwards test (p=0.36) or backwards test (p=0.88). The rounded and unadjusted raw mean for the Digit Span Forward test was 11 and did not differ among the three groups (see Table 16.17).

Table 16.17 : Digit Span Forwards – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	11.19	11.28	10.99
Standard Deviation	2.34	2.42	2.46
50th Percentile	11.00	11.00	11.00
Lower Quartile	9.00	10.00	9.00
Upper Quartile	13.00	13.00	13.00

There was no group effect when adjusted for all potential confounders for all exposed (p=0.37), for Program 1 (p=0.23) or for Program 2 (p=0.27) (see Table 16.18). This was unchanged in the reduced and robust standard error estimates. There was no dose effect (p=0.32).

Table 16.18 : Digit Span Forwards – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	0.22	-0.13	0.58	0.08	2	0.3719
	Richmond	0.18	-0.13	0.50			
Program 1	Amberley	0.36	-0.08	0.81	0.08	2	0.2303
	Richmond	0.30	-0.10	0.70			
Program 2	Amberley	0.35	-0.11	0.82	0.08	2	0.2663
	Richmond	0.32	-0.11	0.75			
Dose	Mild exposure	-0.08	-0.50	0.35	0.08	3	0.3171
	Moderate exposure	-0.36	-0.77	0.06			
	Prolonged exposure	-0.25	-0.66	0.16			

The rounded and unadjusted raw mean for the digit span backwards test was slightly higher for the Richmond group at 7.5 compared with 7.2 for both Amberley and exposed (see Table 16.19).

Table 16.19 : Digit Span Backwards – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	7.20	7.52	7.19
Standard Deviation	2.33	2.35	2.37
50th Percentile	7.00	7.00	7.00
Lower Quartile	5.00	6.00	6.00
Upper Quartile	9.00	9.00	9.00

Similarly, there was no group effect when adjusted for all potential confounders for all exposed (p=0.16), for Program 1 (p=0.09) or for Program 2 (p=0.38) (see Table 16.20). There was no dose effect (p=0.27).

Table 16.20 : Digit Span Backwards – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	0.09	-0.26	0.44	0.08	2	0.1563
	Richmond	0.30	-0.01	0.60			
Program 1	Amberley	0.20	-0.22	0.63	0.08	2	0.0882
	Richmond	0.42	0.04	0.81			
Program 2	Amberley	-0.00	-0.45	0.45	0.07	2	0.3797
	Richmond	0.22	-0.19	0.64			
Dose	Mild exposure	-0.29	-0.70	0.13	0.09	3	0.2691
	Moderate exposure	-0.35	-0.75	0.06			
	Prolonged exposure	-0.16	-0.56	0.24			

16.5.5 New learning/memory

Of the 1538 health study participants eligible for analysis, there were nine who did not complete the Incidental Learning Pairing and Free Recall tests. An additional 17 were excluded due to low score on the Rey and one because of previous stroke. Consequently there were 1511 records available for analysis.

WAIS III Incidental Learning Free Recall

There was no evidence to suggest scores for groups varied across HSA centres (Breslow-Day test for homogeneity of odds ratios: p=0.74 for Richmond versus exposed and p=0.47 for Amberley versus exposed). The exposed group scored in-between the Amberley and Richmond groups; however, the distribution of free recall test scores was far from normal, with the mode, upper quartile and maximum scores equal to nine and the median equal to eight (see Table 16.21). Consequently scores were dichotomised as previously described.

Table 16.21 : Incidental Learning Free Recall – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	7.60	7.91	7.70
Standard Deviation	1.20	1.15	1.27
50th Percentile	8.00	8.00	8.00
Lower Quartile	7.00	7.00	7.00
Upper Quartile	9.00	9.00	9.00

There were 28% of the Amberley group, 21% of the Richmond group and 25% of the exposed in the lower 10th percentile (Table 16.22). Superficially, there might be an expectation that a split at the 10th percentile using the Richmond group would identify an approximate 10% of people in the Richmond group being at or below the 10th percentile. While this might often be the case, it is not always so, depending on the number of similar scores around the 10th percentile. In the present case, the score of 7 was identified as the 10th percentile score for people in the age categories between 40 and 54 years, which includes 57% of the sample, and the score of 7 or below on the whole sample includes 21% of results. The result of 21% in the Richmond group being at the 10th percentile occurs because of a limited range of scores in the tail of the distribution. Another way of explaining

this is that the best score for the bottom 10% of the group also happened to be the same score as the next 11%, hence 21% appear to be in the bottom "10%". Interestingly, a split at the 5th percentile contained 18% from Amberley, 13% from Richmond and 15% from the exposed group (data not shown). Multiple logistic regression revealed no association between group and score classification at or below the 10th percentile for the primary analysis including all exposed (p=0.20), nor for either of the secondary analyses including exposed from Program 1 (p=0.36) or from Program 2 (p=0.33). There was no evidence of a dose association (p=0.75) (see Table 16.23). When analysed by multiple logistic regression with the split at the 5th percentile, there was no change in inference for all exposed, p=0.31 with odds ratios of 1.31 (95% CI: 0.85, 2.0) for Amberley compared to exposed, and 0.94 (95% CI: 0.63, 1.40) for Richmond compared to exposed. These odds differ little from those presented in Table 16.23. Results were similar for the reduced model and with the robust standard variance estimates.

Table 16.22 : Incidental Learning Free Recall – Unadjusted proportions of positive responses from the different groups

	Group			
Free Recall Result*	**Amberley**	**Richmond**	**Exposed**	**Total**
FR <= 10% **≤ 10th percentile of Richmond group**	112 28.07	109 21.37	148 24.58	369
FR > 10% **> 10th percentile of Richmond group**	287 71.93	401 78.63	454 75.42	1142
Total	399	510	602	1511

Frequency missing = 3

** Richmond 10th percentiles selected within age categories*

Table 16.23 : Incidental Learning Free Recall – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	1.23	0.86	1.78	2	3.18	0.2034
	Richmond vs Exposed	0.88	0.63	1.22			
Program 1	Amberley vs Exposed	1.25	0.78	1.99	2	2.07	0.3552
	Richmond vs Exposed	0.94	0.61	1.45			
Program 2	Amberley vs Exposed	1.23	0.77	1.96	2	2.20	0.3325
	Richmond vs Exposed	0.92	0.59	1.43			
Dose	Mild exposure vs Unexposed	1.00	0.64	1.55	3	1.22	0.7493
	Moderate exposure vs Unexposed	0.86	0.55	1.34			
	Prolonged exposure vs Unexposed	1.16	0.76	1.76			

WAIS III Incidental Learning Pairing

There was no evidence to suggest scores for groups varied across HSA centres (Breslow-Day test for homogeneity of odds ratios: p=0.06 for Richmond versus exposed and p=0.78 for Amberley versus exposed). Again, the exposed group scores fell between those of the Amberley and Richmond groups. The distribution of pairing test scores was far from normal, with mode and maximum scores equal to 18 (Table 16.24). Additionally the upper quartile values for both Richmond and exposed were also equal to the maximum possible score. Scores were dichotomised as previously described.

Table 16.24 : Incidental Learning Pairing – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	13.07	14.17	13.38
Standard Deviation	4.47	3.95	4.35
50th Percentile	14.00	16.00	14.00
Lower Quartile	10.00	12.00	10.00
Upper Quartile	16.00	18.00	18.00

There were 24% of the Amberley group, 14% of the Richmond group and 21% of the exposed group in the lower 10th percentile (Table 16.25).

Table 16.25 : Incidental Learning Pairing – Unadjusted proportions of positive responses from the different groups

	Group			
Pairing Result*	**Amberley**	**Richmond**	**Exposed**	**Total**
PR <= 10% **≤ 10th percentile of Richmond group**	94 23.56	71 13.89	128 21.30	293
PR > 10% **> 10th percentile of Richmond group**	305 76.44	440 86.11	473 78.70	1218
Total	399	511	601	1511

Frequency Missing = 3

** Richmond 10th percentiles selected within age categories*

Group was not significantly associated with score classification in any of the multiple logistic regression: p=0.10, 0.17, 0.15 and 0.25 for all exposed, Program 1, Program 2 and dose respectively (Table 16.26). However, the odds of being in the lowest 10th percentile for Richmond versus the exposed group almost reached statistical significance for all exposed (OR 0.21; 95% CI: 0.50, 1.04) and for Program 2 (OR 0.64; 95% CI: 0.41, 1.06).

Table 16.26 : Incidental Learning Pairing – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	1.08	0.74	1.60	2	4.55	0.1028
	Richmond vs Exposed	0.72	0.50	1.04			
Program 1	Amberley vs Exposed	1.38	0.84	2.27	2	3.52	0.1718
	Richmond vs Exposed	0.93	0.58	1.48			
Program 2	Amberley vs Exposed	0.91	0.55	1.49	2	3.80	0.1499
	Richmond vs Exposed	0.66	0.41	1.06			
Dose	Mild exposure vs Unexposed	1.54	0.98	2.41	3	4.10	0.2505
	Moderate exposure vs Unexposed	1.08	0.68	1.71			
	Prolonged exposure vs Unexposed	0.94	0.59	1.50			

Visual Reproduction Immediate Recall

Of the 1538 health study participants eligible for analysis, there were 13 who did not complete at least one of the WMS visual reproduction immediate recall tests. An additional 17 were excluded due to low score on the Rey and one due to previous stroke. Consequently there were 1507 records available for analysis. The distribution of immediate recall test scores was far from normal, with mode and maximum scores equal to 17 (Table 16.27). Additionally the upper quartile value for the exposed was also equal to the maximum possible score. Scores were dichotomised as previously described. There was no evidence to suggest scores for groups varied across HSA centres (Breslow-Day test for homogeneity of odds ratios: $\chi^2 = 4.9$, df = 7, p=0.64 for Richmond versus exposed and $\chi^2 = 4.62$, df = 7, p=0.51 for Amberley versus exposed).

Table 16.27 : Visual Reproduction Immediate Recall – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	13.69	14.46	14.46
Standard Deviation	2.81	2.31	2.53
50th Percentile	14.00	15.00	15.00
Lower Quartile	12.00	13.00	13.00
Upper Quartile	16.00	16.00	17.00

Overall, there was a much larger proportion of Amberley participants in the 10th percentile group at 22% than either Richmond at 14% or exposed at 14% (see Table 16.28).

Table 16.28 : Visual Reproduction Immediate Recall – Unadjusted proportions of positive responses from the different groups

	Group			
Immediate Recall*	**Amberley**	**Richmond**	**Exposed**	**Total**
IR <= 10% **≤ 10th percentile of Richmond group**	88 22.06	69 13.56	85 14.19	242
IR > 10% **> 10th percentile of Richmond group**	311 77.94	440 86.44	514 85.81	1265
Total	399	509	599	1507

Frequency missing = 7

** Richmond 10th percentiles selected within age categories*

Group was significantly associated with score classification for all exposed (p=0.038), Program 1 (p=0.036) and Program 2 (p=0.02) (see Table 16.29).

Table 16.29 : Visual Reproduction Immediate Recall – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	1.17	0.77	1.77	2	6.57	0.0375
	Richmond vs Exposed	0.68	0.46	1.02			
Program 1	Amberley vs Exposed	1.14	0.68	1.92	2	6.66	0.0359
	Richmond vs Exposed	0.66	0.40	1.08			
Program 2	Amberley vs Exposed	1.19	0.68	2.07	2	7.81	0.0201
	Richmond vs Exposed	0.64	0.37	1.11			
Dose	Mild exposure vs Unexposed	0.68	0.37	1.25	3	5.15	0.1611
	Moderate exposure vs Unexposed	1.51	0.92	2.49			
	Prolonged exposure vs Unexposed	1.14	0.68	1.93			

Most of this significance was, again, due to the Amberley versus Richmond comparison, although the comparison between exposed and Richmond groups was on the borderline of significance for all exposed (OR 0.68; 95% CI: 0.46, 1.02), Program 1 (OR 0.67; 95% CI: 0.40, 1.08) and Program 2 (OR 0.64; 95% CI: 0.37, 1.11). This remained on the borderline for the reduced model and became significant for the robust standard error estimates (0.68; 95% CI: 0.5, 0.96). There was no evidence for a dose-response effect (p=0.20).

Visual Reproduction Delayed Recall

Of the 1538 health study participants eligible for analysis, there were 41 who did not complete at least one of the WMS Visual Reproduction Delayed Recall tests. An additional 17 were excluded due to low score on the Rey and one due to previous stroke. Consequently there were 1479 records available for analysis. The distribution of delayed recall test scores was far from normal, with mode and maximum scores equal to 17. Additionally, 16% of

participants obtained the maximum score, and the upper quartile value for the exposed was one point less than the maximum possible score, while the median for the exposed was four points less than the maximum (Table 16.30). Scores were dichotomised as previously described. There was no evidence to suggest scores for groups varied across HSA centres (Breslow-Day test for homogeneity of odds ratios: p=0.79 for Richmond versus exposed and p=0.39 for Amberley versus exposed). Proportionally there were more from the Amberley comparison group in the 10th percentile group at 21% compared to both Richmond at 14% and exposed at 14% (Table 16.31).

Table 16.30 : Visual Reproduction Delayed Recall – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	11.69	12.32	12.37
Standard Deviation	4.20	3.65	3.82
50th Percentile	12.00	13.00	13.00
Lower Quartile	9.00	10.00	10.00
Upper Quartile	15.00	15.00	16.00

Multiple logistic regression provided no evidence of association between group and delayed recall test scores in the 10th percentile for all exposed (p=0.79) Program 1 (p=0.43), or Program 2 (p=0.77). There was no evidence of dose-response (p=0.64) (see Table 16.32). This was also consistent in the reduced model and with the robust standard error estimates.

Table 16.31 : Visual Reproduction Delayed Recall – Unadjusted proportions of positive responses from the different groups

Delayed Recall*	Group			Total
	Amberley	Richmond	Exposed	
DR <= 10% **≤ 10th percentile of Richmond group**	80 20.62	71 14.03	85 14.48	236
DR > 10% **> 10th percentile of Richmond group**	308 79.38	435 85.97	502 85.52	1245
Total	388	506	587	1481

Frequency missing = 33

** Richmond 10th percentiles selected within age categories*

Table 16.32 : Visual Reproduction Delayed Recall – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	1.09	0.71	1.67	2	0.47	0.7902
	Richmond vs Exposed	0.93	0.63	1.39			
Program 1	Amberley vs Exposed	1.45	0.82	2.56	2	1.66	0.4370
	Richmond vs Exposed	1.31	0.77	2.23			
Program 2	Amberley vs Exposed	1.02	0.59	1.77	2	0.53	0.7661
	Richmond vs Exposed	0.88	0.52	1.48			
Dose	Mild exposure vs Unexposed	0.76	0.42	1.37	3	1.71	0.6353
	Moderate exposure vs Unexposed	1.20	0.72	1.99			
	Prolonged exposure vs Unexposed	0.91	0.54	1.54			

Auditory Verbal Learning Test (AVLT)

There were 25 participants in the health study who did not complete the AVLT. An additional 17 participants were removed due to a poor result on the Rey, and a further individual was excluded because of a previous stroke. Consequently, there were 1495 records available for analysis.

Immediate Recall

The average number of words recalled by the exposed group was 9.0 compared with 9.1 from Amberley and 9.6 from Richmond (Table 16.33).

Table 16.33 : AVLT Immediate Recall – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	9.07	9.58	9.00
Standard Deviation	2.83	2.79	2.65
50th Percentile	9.00	10.00	9.00
Lower Quartile	7.00	8.00	7.00
Upper Quartile	11.00	12.00	11.00

There was a significant association between group and number of words recalled for all exposed (p=0.006), Program 1 (p=0.02) and Program 2 (p=0.03) (see Table 16.34). The differences between the exposed group and Amberley were not significant for any of the three analyses, with the 95% confidence interval including zero. There were, however, noticeable differences between the number of words on average recalled by the Richmond group compared to exposed, where the Richmond group scored on average 0.51 points higher (95% CI: 0.16, 0.86) than the exposed for all exposed, on average 0.47 points higher (95% CI: 0.02, 0.91) than the exposed for Program 1, and on average 0.41 points higher (95% CI: -0.08, 0.90) than the exposed for Program 2. This remained significant in the reduced model and increased in significance with the robust standard error estimates. There was no dose response (p=0.16).

Table 16.34 : AVLT Immediate Recall – Summary of multiple linear regression for all exposed Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	-0.02	-0.42	0.38	0.11	2	0.0060
	Richmond	0.51	0.16	0.86			
Program 1	Amberley	-0.04	-0.53	0.46	0.12	2	0.0244
	Richmond	0.47	0.02	0.91			
Program 2	Amberley	-0.12	-0.66	0.41	0.10	2	0.0342
	Richmond	0.41	-0.08	0.90			
Dose	Mild exposure	-0.44	-0.92	0.04	0.11	3	0.1553
	Moderate exposure	-0.21	-0.68	0.26			
	Prolonged exposure	-0.41	-0.88	0.05			

Delayed Recall

The average number of words recalled after delay by the exposed group was 8.8 compared to 9.0 by Amberley and 9.4 by Richmond (Table 16.35).

Table 16.35 : AVLT Delayed Recall – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	9.01	9.38	8.77
Standard Deviation	2.88	2.93	2.89
50th Percentile	9.00	9.00	9.00
Lower Quartile	7.00	7.00	7.00
Upper Quartile	11.00	12.00	11.00

There was a significant association between group and average number of words recalled after delay for all exposed (p=0.03), but not for Program 1 (p=0.18) or Program 2 (p=0.18). The differences between the exposed group and Amberley were not significant for any of the three analyses, with the 95% confidence interval including zero. However, Richmond did

score higher on average, recalling 0.51 (95% CI: 0.14, 0.88) more words than the exposed for all exposed (Table 16.36). This was consistent in the reduced model and with the robust standard error estimates. There was no difference on average between Richmond and exposed for the analysis including Program 1 or Program 2. There was no dose-response effect.

Table 16.36 : AVLT Delayed Recall – Summary of multiple linear regression for all exposed Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	0.22	-0.20	0.65	0.11	2	0.0277
	Richmond	0.51	0.14	0.88			
Program 1	Amberley	0.09	-0.43	0.61	0.12	2	0.1777
	Richmond	0.39	-0.07	0.86			
Program 2	Amberley	0.10	-0.46	0.66	0.10	2	0.1824
	Richmond	0.42	-0.09	0.94			
Dose	Mild exposure	-0.56	-1.07	-0.06	0.11	3	0.0830
	Moderate exposure	-0.25	-0.74	0.24			
	Prolonged exposure	-0.45	-0.94	0.03			

Total Learning

The average total number of words recalled by the exposed group was 44 compared to 45 recalled by Amberley group and 46 recalled by Richmond group (Table 16.37). There was a significant association between group and total number of words recalled for all exposed (p=0.03) and Program 1 (p=0.04) but not Program 2 (p=0.30) (see Table 16.38). The differences between means for the exposed group and Amberley were not significant for any of the three analyses, with the 95% confidence interval clearly including zero. The Richmond group scored on average 1.4 points higher (95% CI: 0.36, 2.51) than the exposed for all exposed, and this was consistent for the reduced model and with the robust standard error estimates. Richmond was on average 1.8 points higher (95%CI: 0.42, 3.14) than the exposed for Program 1, but there was no difference detected between Richmond and exposed for Program 2 as the 95% confidence intervals clearly contained zero. There was no clear dose-response effect (p=0.08).

Table 16.37 : AVLT Total Learning – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	45.24	46.01	44.49
Standard Deviation	8.69	8.60	8.19
50th Percentile	45.00	46.00	45.00
Lower Quartile	39.00	40.00	39.00
Upper Quartile	52.00	52.00	50.00

Table 16.38 : AVLT Total Learning – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	0.80	-0.42	2.03	0.14	2	0.0320
	Richmond	1.43	0.36	2.51			
Program 1	Amberley	1.04	-0.47	2.56	0.16	2	0.0364
	Richmond	1.78	0.42	3.14			
Program 2	Amberley	0.20	-1.40	1.80	0.13	2	0.3037
	Richmond	0.99	-0.48	2.46			
Dose	Mild exposure	-1.24	-2.69	0.22	0.15	3	0.0797
	Moderate exposure	-1.01	-2.43	0.41			
	Prolonged exposure	-1.58	-2.98	-0.18			

Forgetting

The average number of words forgotten was approximately one for Richmond, Amberley and exposed (Table 16.39). There was no statistically significant effect of group in the all exposed comparison or in the Program 1 or Program 2 subgroup comparisons. This was consistent in the reduced model and with the robust standard error estimates. There was no dose-response (p=0.72) (see Table 16.40).

Table 16.39 : AVLT Forgetting – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	1.02	0.99	0.99
Standard Deviation	0.26	0.19	0.21
50th Percentile	1.00	1.00	1.00
Lower Quartile	0.88	0.88	0.88
Upper Quartile	1.13	1.09	1.10

Table 16.40 : AVLT Forgetting – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	0.03	0.00	0.07	0.03	2	0.0621
	Richmond	-0.00	-0.03	0.02			
Program 1	Amberley	0.02	-0.02	0.06	0.04	2	0.2025
	Richmond	-0.01	-0.05	0.02			
Program 2	Amberley	0.03	-0.01	0.08	0.04	2	0.1487
	Richmond	-0.00	-0.04	0.04			
Dose	Mild exposure	-0.02	-0.06	0.02	0.03	3	0.7178
	Moderate exposure	0.00	-0.03	0.04			
	Prolonged exposure	-0.01	-0.05	0.03			

Retrieval Efficiency

Retrieval efficiency was highest for Amberley at 0.77, compared to Richmond at 0.76 and the exposed at 0.75 (Table 16.41). There were no differences between the groups when the multi-variable models were fitted for all exposed (p=0.07), Program 1 (p=0.26) or Program 2 (p=0.56). There was no dose-response (p=0.32) (see Table 16.42). Inferences remained the same in the reduced model and with the robust standard error estimates.

Table 16.41 : AVLT Retrieval Efficiency – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	0.77	0.76	0.75
Standard Deviation	0.21	0.19	0.24
50th Percentile	0.77	0.75	0.75
Lower Quartile	0.64	0.62	0.60
Upper Quartile	0.91	0.92	0.87

Table 16.42 : AVLT Retrieval Efficiency – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	0.04	0.01	0.07	0.06	2	0.0681
	Richmond	0.02	-0.01	0.04			
Program 1	Amberley	0.03	-0.01	0.07	0.07	2	0.2595
	Richmond	0.00	-0.03	0.04			
Program 2	Amberley	0.02	-0.02	0.06	0.05	2	0.5582
	Richmond	0.00	-0.04	0.04			
Dose	Mild exposure	-0.03	-0.07	0.01	0.06	3	0.3213
	Moderate exposure	-0.02	-0.06	0.02			
	Prolonged exposure	-0.02	-0.05	0.02			

16.5.6 Visuospatial ability

1525 participants completed the Block Design test. A further 17 did not pass the Rey, and one had previously suffered a stroke; these were excluded from analyses. Consequently, there were 1507 records available for analysis.

There was no evidence of heterogeneity of group scores between HSA centres (p=0.64); however, centre means differed markedly. The mean score for those tested at Melbourne

was 27, while scores for the other centres ranged between 44.8 and 49.6. The variance for the Melbourne score was 333, compared to a range of 95 to 140 for the scores from the other centres. An examination of variance inequality between groups (using the Bartlett test, pp 614 et seq.)[35] was significant (p=0.02); however, when Melbourne was excluded, there was no evidence to suggest unequal variances between groups (p=0.24). All models described below were fitted with and without data from Melbourne. Parameter estimates shifted slightly but the overall inference did not change; consequently, the models reported below include data from Melbourne HSA centre. Centre was included as a covariate in the multi-variable analysis.

The scores obtained covered the full range of possible values, with two participants scoring zero, 30 participants scoring 12 or less, and five participants scoring the full 68 points. The exposed unadjusted mean raw score of 46 was substantially higher than the mean for Amberley at 42 and slightly lower than the mean for Richmond at 47 (see Table 16.43).

Table 16.43 : Block Design Test – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	42.28	46.85	45.98
Standard Deviation	12.62	12.35	11.20
50th Percentile	44.00	49.00	48.00
Lower Quartile	35.00	41.00	39.00
Upper Quartile	52.00	55.00	54.00

For the multi-variable model including all potential confounders, there was a significant association between group and block design raw score for all exposed (p=0.0001), Program 1 (p<0.0001), and Program 2 (p<0.0001) (see Table 16.44). However, this significance is driven mostly by the difference between Amberley and Richmond comparison groups. When comparing means in the multivariable model including all exposed, Amberley scored on average 2.1 (95% CI: -3.7, -0.5) points lower than the exposed group, and Richmond scored on average 1.6 points (95%CI: 0.2, 2.95) better than the exposed (see group Table 16.44).

Table 16.44 : Block Design Test – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	-2.12	-3.68	-0.56	0.30	2	<.0001
	Richmond	1.60	0.23	2.97			
Program 1	Amberley	-2.39	-4.33	-0.45	0.32	2	<.0001
	Richmond	1.43	-0.31	3.18			
Program 2	Amberley	-2.66	-4.71	-0.60	0.32	2	<.0001
	Richmond	1.35	-0.54	3.25			
Dose	Mild exposure	0.45	-1.39	2.29	0.30	3	0.8267
	Moderate exposure	-0.62	-2.42	1.18			
	Prolonged exposure	0.11	-1.68	1.89			

Bonferroni pair-wise comparison of means revealed highly significant difference between Amberley and Richmond ($p<0.0001$), a difference between Amberley and exposed ($p=0.02$), and no evidence of difference between Richmond and exposed ($p=0.06$). Dichotomising around the 10th percentile for the Richmond group, adjusted for age, and performing multiple logistic regression, showed an overall group effect of 0.002. The odds of being in the Amberley group and scoring in the lowest 10th percentile group were 1.9 (95% CI: 1.2, 3.0) times those of the exposed, while the odds of being in the Richmond group and scoring in lowest 10th percentile group were 0.8 (95% CI: 0.5, 1.3) times those of the exposed. The latter 95% confidence interval clearly contained one, indicating there is no evidence of a significant difference.

When considering exposed from Program 1, the Amberley group scored on average lower than the exposed at -2.4 (95% CI: -4.3, 0.4), and the Richmond group scored on average higher than the exposed at 1.4 (95% CI: -0.3, 3.2), although the latter confidence interval spanned zero. The bulk of the effect observed was due to the difference between the two comparison groups. When considering exposed from Program 2, the Amberley group scored on average lower than the exposed at -2.6 (95% CI: -4.7, -0.6), and the Richmond group scored on average higher at 1.3 (95% CI: -0.6, 3.2), with the latter confidence interval spanning zero, suggesting no difference. As in the previous two models, the bulk of the effect was obtained from the Amberley versus Richmond comparison. There was no significant

difference in mean scores at different levels of dose (p=0.83). The difference with Amberley versus exposed remained significant for the reduced model and with the robust standard error estimates.

16.5.7 Subjective memory complaints

In total, 1725 personnel responded to the Postal Questionnaire. Of these, 18 did not answer all items in the MAC-Q. Consequently there were 1707 complete responses available for analysis. All items on the questionnaire were strongly correlated, with individual correlations between items ranging from 0.55 to 0.72. Single items correlated with all other items very strongly, ranging from 0.70 to 0.85. Cronbach's alpha was 0.88, which was considerably larger than 0.57 reported by Crook et al.,[24] when first describing the questionnaire. It is reasonable to conclude that the MAC-Q used in the current study has strong internal reliability.

The unadjusted mean score was lower for both Richmond and Amberley groups at 23 compared to the exposed group with a mean of 27 (see Table 16.45). Following adjustment for all potential confounders (see Table 16.46), there was a highly significant association between group and self-reported memory loss for the primary analysis and all secondary analyses (p<0.0001 for each analysis).

Table 16.45 : Memory Complaint Questionnaire – Distribution characteristics including means for the three groups

Measure	Amberley	Richmond	Exposed
Mean	22.99	23.36	27.49
Standard Deviation	5.63	5.18	5.54
50th Percentile	23.00	23.00	28.00
Lower Quartile	20.00	21.00	24.00
Upper Quartile	26.00	27.00	32.00

Table 16.46 : Memory Complaint Questionnaire – Summary of multiple linear regression for all exposed, Program 1, Program 2 and Dose

Analysis	Parameter	Estimate	Lower CL	Upper CL	R Square	DF	p-value
All Exposed	Amberley	-4.25	-4.97	-3.53	0.19	2	<.0001
	Richmond	-4.05	-4.67	-3.43			
Program 1	Amberley	-4.78	-5.67	-3.88	0.19	2	<.0001
	Richmond	-4.56	-5.35	-3.77			
Program 2	Amberley	-4.46	-5.39	-3.53	0.16	2	<.0001
	Richmond	-4.15	-5.00	-3.31			
Dose	Mild exposure	3.17	2.29	4.06	0.19	3	<.0001
	Moderate exposure	4.32	3.46	5.17			
	Prolonged exposure	4.81	3.98	5.64			

In the primary analysis where all exposed were included, the Amberley comparison group scored on average 4.25 (95% CI: -4.97, -3.53) points lower than the exposed group, and Richmond scored on average 4.05 (95% CI: -4.67, -3.43) points lower than exposed. These results were similar in the reduced model. A similar trend and scale of difference was observed when considering either those in Program 1 or Program 2 compared with the exposed for Amberley (-4.8 and -4.5 respectively) and Richmond (-4.6 and -4.2 respectively). A significant association between dose and MAC-Q scores was observed ($p<0.0001$) and there was an apparent trend. Those with less than nine months' exposure scored on average 3.2 (95% CI: 2.3, 4.1) points higher than the unexposed; those with nine to 29 months' exposure scored on average 4.3 (95% CI: 3.5, 5.2) points higher than the unexposed; and those reporting 30 or more months' exposure scored on average 4.8 (95% CI: 3.98, 5.6) points higher than the unexposed.

16.5.8 Global measure of cognitive function

There were 1538 participants in the General Health and Medical Study. Of these, two did not undertake the Mini Mental Status Examination (MMSE). A further 80 could not be allocated a

final score because at least one result was not recorded.* There were 1456 records available for analysis. Data were categorised into those who scored less than 24 and those who scored 24 or more as per previous description. Only three participants scored less than 24 points and could be classified as having probable cognitive impairment. All three were from the exposed group. No further analyses could be done.

Table 16.47 shows the distribution of scores achieved on the MMSE for the population of 1456, indicating the three individuals who scored less than 24 and their exposure status.

Table 16.47 : Distribution of MMSE scores

MMSE Score	f	cf	Amberley		Richmond		Exposed		Program	Dose (months)
			N	%	N	%	N	%		
19	1	1	0	*0*	0	*0*	1	*0.2*	P2, W	2 (18)
20	1	2	0	*0*	0	*0*	1	*0.2*	P2, W	1 (7)
23	1	3	0	*0*	0	*0*	1	*0.2*	P1	2 (18)
24	3	6	0	*0*	1	*0.2*	2	*0.3*	-	-
25	8	14	3	*0.8*	0	*0*	5	*0.9*	-	-
26	37	51	7	*1.8*	10	*2.0*	20	*3.4*	-	-
27	49	100	14	*3.7*	13	*2.6*	22	*3.8*	-	-
28	172	272	47	*12*	67	*14*	58	*10*	-	-
29	391	663	109	*29*	123	*25*	159	*27*	-	-
30	793	1456	201	*53*	279	*57*	313	*54*	-	-

f = frequency

cf = cumulative frequency

Shaded area indicates MMSE cut-point

* Depending upon the particular MMSE item which had not been scored by the attending HSA clinician (i.e. having the participant follow the instruction “close your eyes”, which can only be scored with the clinician present), no further summing of items was possible by the study team, so that participant could not be allocated a total score out of 30.

16.5.9 Self-reported symptoms

In total 1725 personnel responded to the Postal Questionnaire. Of these, 15 did not indicate whether they had suffered from forgetfulness in the past month. Of those who did respond in the affirmative, 74% of the exposed reported forgetfulness, as compared with 44% of Richmond and 41% of Amberley (see Table 16.48). After accounting for all potential confounders, a highly significant group effect was observed for all exposed, those in Program 1 and those in Program 2 ($p<0.0001$) (see Table 16.49).

Table 16.48 : PQ item "suffered from forgetfulness in the past month" – Unadjusted proportions of positive responses from the different groups

	Group			
PQ item	**Amberley**	**Richmond**	**Exposed**	**Total**
Self-reported Forgetfulness	198 40.66	259 43.97	471 74.29	928
No self-reported Forgetfulness	289 59.34	330 56.03	163 25.71	782
Total	487	589	634	1710

Frequency missing = 5

Table 16.49 : PQ item "suffered from forgetfulness in the past month" – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.23	0.17	0.32	2	123.30	<.0001
	Richmond vs Exposed	0.28	0.21	0.36			
Program 1	Amberley vs Exposed	0.21	0.14	0.31	2	76.86	<.0001
	Richmond vs Exposed	0.25	0.18	0.35			
Program 2	Amberley vs Exposed	0.24	0.16	0.36	2	55.58	<.0001
	Richmond vs Exposed	0.29	0.20	0.42			
Dose	Mild exposure vs Unexposed	2.80	1.94	4.04	3	115.54	<.0001
	Moderate exposure vs Unexposed	4.08	2.80	5.94			
	Prolonged exposure vs Unexposed	4.78	3.28	6.96			

The odds were substantially reduced for both Richmond and Amberley (0.28 and 0.23 respectively) when compared to the exposed; and the 95% confidence intervals were quite narrow and well below one, suggesting a protective effect from being in these groups as compared with being in the exposed group. This was also consistent in the reduced model. There was also a significant dose effect ($p<0.0001$) and a stepwise increase in risk, with those with less than nine months' exposure having odds of 2.8 (95% CI: 1.94, 4.04) of reporting forgetfulness when compared with the unexposed; those with nine to 29 months' exposure having odds of 4.08 (95% CI: 2.8, 5.94); and those with 30 or more months' exposure having odds of 4.8 (95% CI: 3.3, 6.96).

For loss of concentration there were 16 participants with missing data. Of those who did respond, 69% of exposed, 38% of Richmond comparisons and 34% of Amberley comparisons reported a loss of concentration (see Table 16.50).

Table 16.50 : PQ item "suffered from loss of concentration in the past month" – Unadjusted proportions of positive responses from the different groups

PQ item	Group			Total
	Amberley	Richmond	Exposed	
Self-reported Loss of Concentration	164 33.68	221 37.52	437 69.04	822
No self-reported Loss of Concentration	323 66.32	368 62.48	196 30.96	887
Total	487	589	633	1709

Frequency Missing = 6

After adjusting for all potential confounders, self-report of concentration loss showed a significant group effect for all exposed, those in Program 1 and those in Program 2 ($p<0.0001$) (see Table 16.51). The odds of being in one of the comparison groups and reporting loss of concentration were 0.2 to 0.3 times the odds of being exposed and reporting loss of concentration. The 95% confidence intervals were reasonably narrow and all were well below one; this was also consistent in the reduced model. Moreover, there was a significant dose effect ($p<0.0001$) where the odds of having been exposed for less than nine months and reporting loss of concentration were 3.2 (95% CI: 2.21, 4.55) times the odds of the unexposed; the odds of having been exposed for 9 to 29 months and reporting loss of concentration were 3.8 (95% CI: 2.67, 5.48) times the odds of the unexposed; and the odds of having been exposed for 30 months or more and reporting concentration loss were 4.3 (95%CI: 3.05, 6.12) times the odds of the unexposed.

Table 16.51 : PQ item "suffered from loss of concentration in the past month" – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.23	0.17	0.31	2	129.18	<.0001
	Richmond vs Exposed	0.28	0.22	0.37			
Program 1	Amberley vs Exposed	0.24	0.16	0.34	2	68.94	<.0001
	Richmond vs Exposed	0.29	0.21	0.40			
Program 2	Amberley vs Exposed	0.26	0.18	0.38	2	50.91	<.0001
	Richmond vs Exposed	0.33	0.24	0.47			
Dose	Mild exposure vs Unexposed	3.17	2.21	4.55	3	119.25	<.0001
	Moderate exposure vs Unexposed	3.83	2.67	5.48			
	Prolonged exposure vs Unexposed	4.32	3.05	6.12			

There were 15 participants who did not indicate whether they had suffered from difficulty finding the right word in the past month or not. As shown in Table 16.52, 71% of the exposed, 45% of the Richmond and 42% of the Amberley comparisons reported difficulty finding the right word. After adjusting for all potential confounders, self-reported difficulty finding the right word showed a significant group effect for all exposed, those in Program 1 and those in Program 2 ($p<0.0001$). The odds of being in one of the comparison groups and reporting difficulty finding the right word were 0.3 to 0.4 times the odds of being exposed and reporting difficulty finding the right word (Table 16.53).

Table 16.52 : PQ item "suffered from difficulty finding the right word in the past month" – Unadjusted proportions of positive responses from the different groups

PQ item	Group			Total
	Amberley	Richmond	Exposed	
Self-reported Difficulty finding right word	203 41.68	267 45.33	447 70.50	917
No self-reported Difficulty finding right word	284 58.32	322 54.67	187 29.50	793
Total	487	589	634	1710

Frequency Missing = 5

The 95% confidence intervals were reasonably narrow and all were well below one; this was consistent in the reduced model as well. Moreover, there was a significant dose effect ($p<0.0001$), but the increase over levels of dose was not linear. The odds of having been exposed for less than nine months and reporting difficulty finding the right word were 2.4 (95% CI: 1.68, 3.44) times the odds of the unexposed; the odds of having been exposed for nine to 29 months and reporting difficulty finding the right word were 3.4 (95% CI: 2.32, 4.83) times the odds of the unexposed; and the odds of having been exposed for 30 months or more and reporting difficulty finding the right word were 3.2 (95% CI: 2.3, 4.6) times the odds of the unexposed.

Table 16.53 : PQ item “suffered from difficulty finding the right word in the past month” – Summary of multiple logistic regression for all exposed, Program 1, Program 2 and Dose

Analysis	Effect	Odds Ratio	Lower CL	Upper CL	DF	Wald Chi Square	p-value
All Exposed	Amberley vs Exposed	0.29	0.22	0.39	2	90.37	<.0001
	Richmond vs Exposed	0.35	0.28	0.46			
Program 1	Amberley vs Exposed	0.26	0.18	0.37	2	59.75	<.0001
	Richmond vs Exposed	0.31	0.22	0.43			
Program 2	Amberley vs Exposed	0.32	0.22	0.47	2	36.41	<.0001
	Richmond vs Exposed	0.40	0.28	0.57			
Dose	Mild exposure vs Unexposed	2.40	1.68	3.44	3	82.13	<.0001
	Moderate exposure vs Unexposed	3.35	2.32	4.83			
	Prolonged exposure vs Unexposed	3.23	2.28	4.58			

16.6 Discussion

Executive functioning

Of the four objective tests used to assess this domain, two indicate a poorer score for the F-111 DSRS group compared to Richmond and Amberley comparisons. The COWAT Letter and Animal tasks are both significant in the overall comparison, and there is a clear dose-response curve, but this significance is lost in the Program 1 and 2 subgroups and reduced with the robust stand error estimates. Surprisingly, the Trail Making Test Part B indicated a poorer score for the Richmond group compared to the Amberley comparisons or exposed group.

To put these results in context, normative values were obtained for our population, i.e. average 44-year-old male with high school or university education. For the COWAT Letter task, average scores put the comparison groups at approximately the 40% percentile and the exposed group at the 30% percentile; for the COWAT Animal task, the comparison groups are roughly in the 50% percentile while the exposed are in the 40% percentile. Hence these results indicate a drop in executive function in the DSRS group equivalent to a decile drop in normative scores. These results are tempered by the lack of any effect in the exposed group for the other two tests in this domain: WAIS III Similarities and Trail Making Test Part B.

Psychomotor speed

All three tests of psychomotor speed indicated a statistically significant decrease in performance in the DSRS group, especially compared to the Richmond group. All three tests consistently show a statistically significant or borderline significant decrease for all exposed, as well as in Program 1 and 2 subgroups compared to the Richmond comparison group. There was a suggestion of a dose-response effect for only one of the tests: Digit Symbol Coding.

The magnitude of the decrease in performance appears to be quite small. For example, the mean scores for the Trail Making Test Part A indicate that despite the significant difference, both the exposed and comparison groups are scoring around the 70% percentile, and the difference is clinically small at approximately one second.

Attention/working memory

Attention/working memory was assessed using the Forward and Backward Digit Span tests. Neither of these tests indicated any difference between the comparison and exposed groups; the absence of significant results was consistent in the overall comparisons, Program 1 and 2 subgroup comparisons, and there was no evidence of a dose-response effect. This may indicate a true absence of any effect, or may be due to the poor sensitivity of this test to pick up subtle changes in this domain.

New learning/memory

The three tests administered to assess the ability to learn new material were: (a) Incidental Learning Free Recall and Incidental Learning Pairing, (b) Visual Reproduction Immediate and Delayed Recall, and (c) Auditory Verbal Learning Test (with five sections: immediate recall, delayed recall, total learning, forgetting, and retrieval efficiency). The first two tests – digit symbol and visual reproduction – proved too insensitive in our population. Most participants, regardless of group, scored at or near the maximum result, a phenomenon termed a “ceiling effect”. Hence these tests did not have enough ability to discriminate or detect subtle changes between participants. There were no consistent and convincing differences between the exposed and comparison groups, although some comparisons between the exposed group and Richmond group were of borderline significance, e.g. Incidental Learning Pairing and Immediate Recall on Visual Reproduction.

The AVLT performed much better, giving a good spread of data for all three groups, as evidenced by a normal distribution. For all five sections, the pattern of results is generally similar, with the exposed group doing slightly worse than both the Amberley and Richmond comparisons. However, only three of the five sections’ results were significant (immediate recall, delayed recall, and total learning), with the exposed group doing worse than the Richmond comparisons. In the immediate and delayed recall test, this decrement in performance was equivalent to about half a point worse (out of 18), and in the total learning it was 1.4 points worse; both of these equate to about half to three-quartes of a decile change in performance. These differences remained statistically significant or close to significant in both Program 1 and 2 subgroup analyses. However, none of these gave a clear dose-response effect. The clinical impact of such a difference is probably very small. Setting a clinical threshold at the lowest 10^{th} percentile led to the finding of no statistical difference between groups on any of these tests. The other two sections of the AVLT, forgetting and retrieval, did not detect any difference between exposed and comparison groups.

Visuospatial ability

The WAIS III Block Design Test measured visuospatial ability, with higher scores indicating better performance. The overall test of significance across all three groups was strongly positive, but this was driven mainly by the comparison between Amberley and Richmond groups. The exposed group performed slightly better than the Amberley group, but slightly worse than the Richmond group, by approximately 1.5 points. This point estimate was consistent for Program 1 and 2 subgroups, but there was no dose-response effect seen. The magnitude of this difference is approximately equivalent to half a decile change in score and is unlikely to be clinically significant. Indeed, when analysed using a clinically significant threshold of the 10th percentile, this difference was not statistically significant.

Subjective memory complaints

The MAC-Q is a validated questionnaire for measuring self-reported memory complaints (subjective memory). The MAC-Q proved to be a particularly strong and reliable tool in the current study, with internal validity scores even higher than those reported previously. There was a strong, consistent and statistically significant increase in memory complaints in the exposed group relative to both the Amberley and Richmond comparison groups. This increase was also shown in Program 1 and 2 subgroups and showed a dose-response effect, increasing approximately 3, 4 and 5 points in the “mild”, “moderate” and “prolonged” exposure groups respectively relative to the non-exposed group. The increase of approximately 4.5 points between the exposed and comparison groups is reasonably large when judged in the context of the total test score of 35; this is equivalent to an approximately two-decile increase in memory complaints and is likely to be clinically meaningful. Taken as a dichotomous measure (i.e. memory complaints present or absent) there is a consistent and statistically significant four-fold increase in memory complaints in the exposed group compared to either comparison group.

Global measure of cognition

With regard to the MMSE, there were too few individuals with scores less than the threshold to do any analysis. We conclude that the MMSE, being a general screening test of cognitive ability, may be too general and blunt to pick up any possible subtle neurocognitive changes that may be present in the current study’s population. In essence, most individuals in all three groups scored at, or close to, a perfect score: another example of a “ceiling effect”, i.e. the test was not discriminatory enough. This result supports the decision to include a battery of specific cognitive tests tailored to particular domains as more sensitive measures of neuropsychological impairment.

Self-reported symptoms

There was an increase in self-reported symptoms of forgetfulness, loss of concentration, and difficulty in finding the right word; the prevalence of these symptoms is between 70-75% in the exposed group and between 35-45% in the comparisons. This corresponds to a statistically significant increase of 2.8-4.3 fold in the occurrence of these symptoms; all three symptoms are consistent in showing an increase relative to the Amberley and Richmond comparison groups, an increase in Program 1 and 2 subgroups and a dose-response effect.

Impact of depression and anxiety on cognition

It is interesting to note that in many of the multivariate regression analyses in this chapter, depression or anxiety were significantly associated with cognitive outcomes. It is well established that anxiety and depression can impair cognitive performance.[33,34] When exploring the models with and without these mood variables, it was apparent that while anxiety and depression explained some of the variance in the outcomes, they did not alter the point estimates for group to any appreciable degree. This means that the group effect on cognition seen with DSRS activities cannot be attributed solely to the concomitant mood disorder. In other words, the association between DSRS and cognitive outcomes is at least partly independent of the effect of mood; this raises the possibility that these apparent cognitive decrements, though subtle or small, may be “organic” or “physiological” in origin rather than mediated through a mechanism of depression or anxiety.

16.7 Conclusions

Subjective assessments of cognition showed a strong, significant and graded decrease in function in the exposed group relative to the two comparison groups, whether by questionnaire or by using a validated measure such as the MAC-Q. This is supported to some extent by the objective measures. Two of the tests in the executive functioning domain showed decreased performance for the normative scores. In the new learning/memory domain, two tests raise the possibility of decreased function in the exposed group versus Richmond only, and three out of five sections on the third test (AVLT) confirmed this with statistically significant results. For psychomotor speed all three tests showed a consistent, statistically significant decrease in function for the DSRS group compared to the Richmond group. There were no detectable differences in attention/working memory or visuospatial ability.

In summary, there is a strong and consistent increase in self-reported cognitive problems among the exposed. This is supported by the objective tests to some degree, in that three of the five domains within cognition (executive functioning, new learning/memory, and psychomotor speed) consistently show poorer performance for the F-111 DSRS group versus the Richmond comparisons. This effect is independent of any mood disorder (i.e. depression or anxiety).

16.8 References

1. Donaldson, E. Nature and extent of health complaints, Report to the F111 Deseal/Reseal Board of Inquiry. Volume 2, Part 1, Chapter 13, Annex B. (RAAF, Amberley, 2000).
2. Filley, C.M. & Kelly, J.P. Chapter 19, Clinical neurotoxicity and neurobehavioural toxicology. in *Clinical Environmental Health and Toxic Exposures* (ed. Krieger, G.R.) 247-259 (Lippincott Williams & Wilkins, Philadelphia, 2001).
3. Anger, W.K. Neurobehavioural testing of chemicals; impact on recommnded standards. *Neurobehavioural Toxicology & Teratology* **6**, 147-153 (1984).
4. Foo, S.C., Jeyaratnam, J. & Koh, D. Chronic neurobehavioural effects of toluene. *British Journal of Industrial Medicine* **47**, 480-484 (1990).
5. Cherry, N., Hutchins, H., Pace, T. & Waldron, H.A. Neurobehavioural effects of repeated occupational exposure to toluene and paint solvents. *British Journal of Industrial Medicine* **42**, 291-300 (1985).
6. Fornazzari, L., Wilkinson, D.A., Kapur, B.M. & Carlen, P.L. Cerebellar, cortical and functional impairment in toluene abusers. *Acta Neurologica Scandinavia* **67**(1983).
7. Morrow, L.A., Ryan, C.M., Hodgson, M.J. & Robin, N. Alterations in cognitive and psychological functioning after organic solvent exposure. *Journal of Occupational Medicine* **32**, 444-50 (1990).
8. Morrow, L.A., Ryan, C.M., Hodgson, M.J. & Robin, N. Risk factors associated with persistence of neuropsychological deficits in persons with organic solvent exposure. *Journal of Nervous and Mental Disease* **179**, 540-5 (1991).
9. Morrow, L.A. & Scott, A. Comparison of neuropsychological test scores between men and women with prior exposure to organic solvents. *Applied Neuropsychology* **9**, 240-3 (2002).
10. Morrow, L.A., Stein, L., Bagovich, G.R., Condray, R. & Scott, A. Neuropsychological assessment, depression and past exposure to organic solvents. *Applied Neuropsychology* **8**, 65-73 (2001).
11. Bolla, K.I. et al. Comparison of neurobehavioral function in workers exposed to a mixture of organic and inorganic lead and in workers exposed to solvents. *American Journal of Industrial Medicine* **27**, 231-246 (1995).
12. Hogstedt, C. Has the Scandinavian solvent syndrome controversy been solved? *Scandinavian Journal of Work and Environmental Health* **20**, 59-64 (1994).
13. Knave, B., Persson, H.E., Goldberg, J.M. & Westerholm, P. Long-term exposure to jet fuel: an investigation on occupationally exposed workers with special reference to the

nervous system. *Scandinavian Journal of Work and Environmental Health* **2**, 152-64 (1976).

14. Tu, R.H., Mitchell, C.S., Kay, G.G. & Risby, T.H. Human exposure to the jet fuel, JP-8. *Aviation, Space and Environmental Medicine* **75**, 49-59 (2004).
15. White, R.F. & Proctor, S.P. Research and clinical criteria for the development of neurobehavioural test batteries. *Journal of Occupational Medicine* **23**, 140-148 (1992).
16. Feldman, R.G., Ratner, M.H. & Ptak, T. Chronic toxic encephalopathy in a painter exposed to mixed solvents. *Environmental Health Perspectives* **107**, 417-422 (1999).
17. Wechsler, D. *Wechsler Adult Intelligence Scale*, (The Psychological Corporation, San Antonio, Texas, 1997).
18. Cardall, A.J. *Purdue Pegboard*, (Science Research Associates, Chicago, USA., 1943).
19. Ryan, C.M., Morrow, L.A., Bromet, E.J. & Parkinson, D.K. Assessment of neuropsychological dysfunction in the workplace: Normative data from the Pittsburgh Occupational Exposures Test Battery. *Journal of Clinical and Experimental Neuropsychology* **9**, 665-679 (1987).
20. Sunderland, A., Harris, J.E. & Baddeley, A.D. Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behaviour* **22**, 341-357 (1983).
21. Sunderland, A., Harris, J.E. & Gleave, J. Memory failures in everyday life following severe head injury. *Journal of Clinical Neuropsychology* **6**, 127-142 (1984).
22. Hanninen, T. et al. Subjective memory complaints and personality traits in normal elderly subjects. *Journal of the American Geriatric Society* **42**, 1-4 (1994).
23. Turvey, C.L., Schultz, S., Arndt, S., Wallace, R.B. & Herzog, R. Memory complaint in a community sample aged 70 and older. *Journal of the American Geriatric Society* **48**, 1435-1441 (2000).
24. Crook, T.H., Feher, E.P. & Larabee, G.J. Assessment of memory complaint in age-associated memory impairment: The MAC-Q. *International Psychogeriatric* **4**, 165-175 (1992).
25. Folstein, M.F., Folstein, S.E. & McHugh, P.R. Mini-Mental State: A practical method for grading the state of patients for the clinican. *Journal of Psychiatric Research* **12**, 189-198 (1975).
26. O'Connor, D.W. et al. The reliability and validity of the Mini-Mental State in a British community survey. *Journal of Psychiatric Research* **23**, 87-96 (1989).
27. Ponds, R.W., Verhey, F.R., Rozendaal, N., Jolles, J. & Deelman, B.G. Screening for dementia: validity of the cognitive screening test and the mini-mental state examination. *Tijdschrift Voor Geroltolgoie En Geriatrie (Deventer)* **23**, 94-99 (1992).

28. Ishizaki, J., Meguro, K. & Ambo, H. A normative, community-based study of mini-mental state in elderly adults: the effect of age and educational level. *The Journals of Gerontology: Psychological Sciences and Social Sciences* **53**, 359-363 (1998).
29. Kurlowicz, L. & Wallace, M. The mini-mental state examination. *Journal of Gerontological Nursing* **25**, 8-9 (1999).
30. Tombaugh, T.N. & McIntyre, N.J. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatric Society* **40**, 922-935 (1992).
31. Antikainen, R. et al. Mood improvement reduces memory complaints in depressed patients. *European Archives of Psychiatry and Clinical Neuroscience (Berlin)* **251**, 6-11 (2001).
32. Tobiansky, R., Blizard, R., Livingston, G. & Mann, A. The Gospel Oak Study stage IV: the clinical relevance of subjective memory impairment in older people. *Psychological Medicine* **25**, 779-786 (1995).
33. La Rue, A., Swan, G.E. & Carmelli, D. Cognition and depression in a cohort of aging men: results from the western collaborative group study. *Psychology and Aging* **10**, 30-33 (1995).
34. Backman, L., Hill, R.D. & Forsell, Y. The influence of depressive symptomatology on episodic memory functioning among clinically nondepressed older adults. *Journal of Abnormal Psychology* **105**, 97-105 (1996).
35. Neter, J., Wasserman, W. & Kutner, M.H. *Applied Linear Statistical Models*, (Irwin Inc., Boston, 1990).

17 Review of Findings and Discussion

Chapter summary

This chapter explores how the various positive findings in the Deseal/Reseal (DSRS) group "cluster" together. This involves a two-step process: firstly, comparing findings for individuals within a domain; secondly, comparing results for individuals across domains. This process identified 11 main outcomes:

a) Four continuous outcomes: SF-36 (quality of life) physical component score, SF-36 mental component score, neuropsychological function, and subjective memory impairment, which were not correlated with each other.

b) Seven dichotomous outcomes: self-reported physician-diagnosed obstructive lung disease, self-reported physician-diagnosed dermatitis, skin lesions including moles/naevi, male sexual function, General Health Questionnaire (GHQ), and depression and anxiety (as measured by the computerised CIDI assessment), of which only the last three were fairly correlated with each other.

There were similar patterns of associations across the dichotomous and continuous measures, indicating internal consistency of the data. While there are unavoidable uncertainties in data interpretation, the results point to an association between DSRS and poor physical and mental quality of life, erectile dysfunction, depression, anxiety, and subjective memory impairment.

Chapter contents

17.1 Introduction

In the chapters thus far, each measure has been discussed in isolation. This chapter seeks to explain how the findings on these individual measures "cluster" together. The aim was two-fold:

a) To compare significant findings for individuals **within** a domain. For example, within the neuropsychological outcomes, were the subjects who performed poorly on executive function the same ones who scored poorly on new learning or psychomotor speed?
b) To compare significant findings for individuals **across** domains. For example, were the subjects who did worse on neuropsychological tests the same ones who had high scores on the depression scale or who had dermatitis detected during the physical exam?

Finally, conclusions are drawn about the overall coherence of the data and the significance of the findings.

17.2 Approach

The general approach was as follows:

1) Results from each domain were summarised in tabular form. Each of these tables corresponds to a chapter of results; and the tables are presented in the same order as the chapters.

2) Within each domain, the principal positive findings were compared with each other, regardless of whether those findings were self-reported symptoms, self-reported previous physician diagnoses, or physical examination findings.

3) From these were identified the main or strongest positive results "representative" of that domain. In order to be as objective as possible, self-reported symptoms were excluded from this step. In most cases, one summary measure was chosen; however, if all positive findings in a domain were self-reported symptoms, then no representative measure was taken from this domain and it was not analysed further. If two or more findings segregated separately, e.g. those who had depression were different from

those who had anxiety, then both measures were chosen as “representative” and used for further analysis.

4) Finally, representative measures across domains were compared in order to see which findings “cluster” together in the DSRS group.

17.3 Analyses and methods

Continuous outcomes were correlated using Spearman’s rank correlation coefficient, which is a non-parametric correlation appropriate for non-normally distributed data. Dichotomous outcomes were compared using the kappa coefficient, which expresses the level of agreement corrected for chance. The interpretation of these coefficients according to general ranges is given in Table 17.1.[1]

Table 17.1 : Level of agreement for kappa coefficients

Level of agreement	Value of coefficient
Weak	<0.4
Fair	0.4-0.59
Moderate	0.6-0.79
Strong	>0.8

Adapted from: Fleiss, J.L. (1981). Statistical methods for rates and proportions. In: Statistical Methods for Rates and Proportions. John Wiley and Sons.

It is important to note that kappa also depends on the prevalence of a finding. For example, let us assume that only a few people reported a physician diagnosis of obstructive lung disease (i.e. bronchitis and emphysema), and many more reported wheezing; even if all those with a physician diagnosis reported wheezing, the kappa will be poor due to all those who reported wheezing without having a physician diagnosis. For this reason the raw numbers for overlap were also reported when using kappa coefficients.

Where we needed to select the poorest result from a series of continuous outcomes measured on different scales (for example, choosing the worst result from the four different tests that measured executive function within the neuropsychological domain), the results

were standardised and expressed as a Z-score. The Z-score transforms scores to a distribution with a mean of 0 and a standard deviation of 1, and hence creates “common units” for comparison across different scales. Some outcomes were reversed so that a low value represented a poor score rather than a good score. The worst score for each individual was then the lowest of these Z-score values. Where a dichotomous measure was to be associated with a continuous measure, the following method was used: the mean values of the continuous measure for each of the two groups on the dichotomous measure were compared. The association was measured by the difference between these means as well as by the significance level, using analysis of variance adjusted for age, rank and posting (see Table 17.2).

Table 17.2 : Illustration of association for dichotomous and continuous outcomes

Dichotomous measure	Continuous measure	
Yes	Mean 1	} Δ (difference)
No	Mean 2	

17.4 Results

17.4.1 Within domain associations

General health and well-being

The results of Chapter 9 (General Health and Well-Being) are summarised in Table 17.3. As for all summary tables for each domain in this section, the columns detail the following information from left to right:

- domain
- the overall magnitude of the association and the significance level; the “OR” (odds ratio) or “diff” (difference) reported is that for the exposed group versus Amberley or Richmond where there was a significant difference, or the average OR/difference is reported where both groups were significantly different
- whether the association is significant (+) or not (-) compared to the Amberley (A) or Richmond (R) group

- whether the results are consistent (+) or not (-) for the Program 1 (P1) and 2 (P2) subgroups, based on the point estimate and the p-value
- whether there is a dose-response effect, as judged by a stepwise increase in the point estimates and a significant overall p-value for dose effect; if neither of these two criteria are met, there is no dose-response effect (N); if only one of the criteria is met, there is an equivocal dose-response (+/-); and if both are met, there is a dose-response effect (Y)
- if the overall association is not significant, then no further results are presented.

Given that the strength of the inference for hospitalisation was weak – i.e. only significant against Richmond, only against Program 2, and no dose-response – this was omitted from the within-domain associations.

Table 17.3 : Summary of general health and well-being results

Domain	Overall comparison	Group comparison	Subgroup comparison	Dose-resp.
a. Hospitalisation	OR 1.5, p<0.05	A– R+	P1– P2+	N
b. SF-36 physical component score	-2.3 diff, p=0.004	A– R+	P1+ P2+	Y
c. SF-36 mental component score	-6.9 diff, p<0.0001	A+ R+	P1+ P2+	Y

The physical and mental component scores of the SF-36 are the findings with the strongest inference within this domain, i.e. both significant in both subgroup comparisons and both showing a dose-response effect. The correlation between these two scores was very weak, with a Spearman correlation coefficient of 0.04. This indicates that those scoring poorly on one scale do not necessarily score poorly on the other, and argues that these are essentially independent, i.e. they represent two distinctly different populations.

This separation between those who score poorly on the physical scale and those who score poorly on the mental scale argues that there is some specificity to exposed participant responses, i.e. people are not just scoring poorly in both domains in a non-specific manner. Hence, both physical and mental component scores were retained for comparison across domains.

Cardiovascular health

The results of Chapter 10 (Cardiovascular Health) are summarised in Table 17.4. There is strong evidence for a 2- to 2.5-fold increase in various self-reported cardiovascular symptoms.

Table 17.4 : Summary of cardiovascular results

Domain	Overall comparison	Group comparison	Subgroup comparison	Dose-resp.
a. Postural hypotension systolic	-0.10 diff, p=0.97	-	-	-
b. Postural hypotension diastolic	-0.10 diff, p=0.96	-	-	-
c. Self-reported (SR) dizziness in the past month	OR 2.5, p<0.0001	A+ R+	P1+ P2+	Y
d. SR feeling faint or fainting when standing	OR 2.5, p<0.0001	A+ R+	P1+ P2+	Y
e. SR chest pain in the past month	OR 2.3, p<0.0001	A+ R+	P1+ P2+	Y
f. SR rapid / pounding / irregular heartbeat	OR 2.0, p=0.0001	A+ R+	P1+ P2+	Y
g. SR physician diagnosed high BP	OR 0.91, p=0.59	-	-	-

The agreement between the various cardiovascular symptoms is shown in Table 17.5. The numbers in the cells of the table represent the number of individuals with a positive outcome for both variables; for example there were 86 exposed individuals who reported chest pain and tachycardia. Surprisingly, there is only weak agreement between these symptoms, with most kappas being less than 0.4. This again demonstrates that there is a reasonably high degree of selectivity for the symptoms reported by the DSRS group.

Table 17.5 : Agreement between self-reported cardiovascular symptoms

	Chest Pain N=167	Tachycardia N=146	Dizziness N=191	Faint when standing N=144
Chest Pain		K=0.40 N=86	K=0.35 N=98	K=0.24 N=68
Tachycardia			K=0.31 N=84	K=0.2 N=57
Dizziness				K=0.39 N=94

Given that all the positive findings are based on self-reported symptoms, these are not compared across domains.

Respiratory health

The results of Chapter 11 (Respiratory Health) are summarised in Table 17.6. Although there were no significant findings on spirometry, there was a significant, two-fold higher incidence of previous physician-diagnosed obstructive lung disease (i.e. bronchitis and emphysema), and a similar increase in self-reported shortness of breath or wheezing. The agreement between these two measures was poor (kappa=0.17), with two-thirds of the group with physician-diagnosed obstructive lung disease reporting wheezing. The more objective measure of physician-diagnosed obstructive lung disease was retained for comparison across domains.

Table 17.6 : Summary of respiratory results

Respiratory Health	Overall comparison	Group comparison	Subgroup comparison	Dose-resp
a. Obstructive airways disease (HE)*	OR 1.3, p=0.47	-	-	-
b. SR physician diagnosed asthma	OR 1.1, p=0.52	-	-	-
c. SR previous physician diagnosed obstructive lung disease	OR 2.0, p<0.0001	A+ R+	P1+ P2+	+ / -
d. SR shortness breath/wheezing in the past month	OR 1.8, p<0.0001	A+ R+	P1+ P2+	Y

** Health Examination – spirometry testing for asthma-like symptoms*

Dermatological and breast abnormalities

The results of Chapter 12 (Dermatological and Breast Abnormalities) are summarised in Table 17.7. The findings with the strongest evidence were the three measures of dermatitis: self-reported rash, previous physician-diagnosed dermatitis, and dermatitis noted at the physical examination. The first two were consistent across comparison groups and programs, and showed some evidence of a dose-response effect; the latter was somewhat weaker, showing significance only against Amberley and only in Program 2.

The agreement coefficients between the three measures of dermatitis are given in Table 17.8 and are generally weak, although this may be due to the transient nature of skin manifestations. The results indicate that less than 100 people had dermatitis at the time of the physical exam, but over half of these reported a previous physician diagnosis of dermatitis. Almost 80% of those with a physician diagnosis also reported a rash, but most of those with a self-reported rash had no formal diagnosis, either previously from a physician or on examination.

Table 17.7 : Summary of dermatological and breast results

Dermatological and breast abnormalities	Overall comparison	Group comparison	Subgroup comparison	Dose-resp.
a. Self-reported skin rash or irritation	OR 2.4, p<0.0001	A+ R+	P1+ P2+	Y
b. Self-reported skin ulcer	OR 1.8, p=0.03	A– R+	P1+ P2+	+ / -
c. Self-reported physician diagnosis of psoriasis	OR 1.3, p=0.31	-	-	-
d. Self-reported physician diagnosis of dermatitis	OR 2.6, p<0.0001	A+ R+	P1+ P2+	+ / -
e. Self-reported physician diagnosis of eczema	OR 1.4, p=0.30	A– R–	P1– P2–	N
f. Self-reported physician diagnosis of melanoma	OR 1.4, p=0.10	A+ R–	P1– P2+	N
g. Health examination diagnosed dermatitis	OR 1.5, p=0.05	A+ R–	P1– P2+-	+ / -
h. Health examination diagnosed eczema	OR 1.2, p=0.62	A– R–	P1– P2–	N
i. Health examination diagnosed melanoma/BCC/SCC	OR 1.2, p=0.62	-	-	-
j. Health examination pigmented or sun-related lesions	OR 1.7, p=0.0009	A+ R+	P1+ P2+	Y

Table 17.8 : Agreement between measures of dermatitis

	SR Rash N=307	Dermatitis (physician diagnosed) N=183	Dermatitis (health exam) N=97
Rash		K=0.32 N=142	K=0.20 N=78
Dermatitis (PQ)			K=0.21 N=50

The other positive finding in this domain was for pigmented or sun-related lesions (e.g. moles/naevi, and actinic/solar keratoses). The association between physician-diagnosed dermatitis and the latter was very poor, with a kappa of –0.02, indicating little overlap in these two outcomes. Given the semi-objective nature of a previous physician diagnosis and the strength of the inference, this was taken as the best measure for comparison across domains.

Neurological outcomes

The results of Chapter 13 (Neurological Outcomes) are summarised in Table 17.9. The three strongest findings are self-reported sensory symptoms, self-reported motor neuropathic symptoms, and self-reported sensitivity to smells, all in the range of a 2- to 2.5-fold increase in odds relative to comparisons. Colour vision was also abnormal when measured using the Colour Confusion Index (CCI) but not when measured with a clinical diagnosis; this lack of consistency between the two measures indicates that the association with colour vision abnormality is weak.

Table 17.10 shows the agreement between the four measures within the exposed group. Sensory neuropathic symptoms are the most common self-reported symptom. Over 80% of those with motor neuropathic symptoms also report sensory neuropathic symptoms (170 out of 203 individuals), and around 50% of those with sensory symptoms also report motor symptoms (170 out of 346 individuals). There is little overlap between colour vision and motor neuropathic symptoms, but more overlap with sensory neuropathic symptoms, i.e. 60% of those with colour vision deficiency also report sensory symptoms (135 out of 219 individuals). There is little overlap between sensitivity to smells and colour vision loss but

over two thirds of those with sensitivity to smells report motor and sensory symptoms (64 out of 90 individuals).

Sensory neuropathic symptoms appear to be the most consistent finding across all those with abnormalities in this area, but given that this is not an objectively-diagnosed finding, comparisons were not taken further, i.e. across domains.

Table 17.9 : Summary of neurological outcomes

Neurological outcomes	Overall comparison	Group comparison	Subgroup comparison	Dose-resp
a. Vibration perception threshold score – finger	0.05 diff, p=0<0.05	A– R+	P1– P2–	N
b. Vibration perception threshold score – toe	0.01 diff, p=0.37	-	-	-
c. Self-reported sensory symptoms	OR 2.0, p<0.0001	A+ R+	P1+ P2+	Y
d. Self-reported motor symptoms	OR 2.6, p<0.0001	A+ R+	P1+ P2+	Y
e. Sniffin' Sticks olfaction score	-0.30 diff, p=0.16	-	-	-
f. Self-reported sensitivity to smells/odours in past month	2.5 diff, p<0.0001	A+ R+	P1+ P2+	Y
g. Colour Confusion Index (CCI) for colour vision	OR 1.3, p=0.06	R+	P1– P2–	+ / -
h. Clinical diagnoses of colour vision deficits	OR 1.3, p=0.11	-	-	-

Table 17.10 : Agreement between neurological outcomes

	Colour vision N=219	Motor N=203	Sensory N=346	Smell sensitivity N=90
Colour vision		K=0.05 N=79	K=0.09 N=135	K=0.02 N=34
Motor			K=0.33 N=170	K=0.29 N=64
Sensory				K=0.13 N=74

Sexual function and reproductive health

The results of Chapter 14 (Sexual Function and Reproductive Health) are summarised in Table 17.11. There was a consistent 2.5-fold increase in male sexual dysfunction, whether assessed by self-report or the validated International Index of Erectile Function (IIEF) questionnaire. As shown in Table 17.12, there was moderately good agreement between these three measures, with 70-80% of those who self-reported problems having poor IIEF results. The IIEF results were thus taken as representative for this domain. The dichotomous scale for erectile function was used rather than the continuous scale for overall function, as the continuous distribution required complex transformations and the parameter estimates were difficult to interpret.

Table 17.11 : Summary of sexual function and reproductive health results

Sexual function	Overall comparison	Group comparison	Subgroup comparison	Dose-resp
a. SR loss of interest in sex	OR 2.5, p<0.0001	A+ R+	P1+ P2+	+ / -
b. SR problems with sexual functioning	OR 2.5, p<0.0001	A+ R+	P1+ P2+	+ / -
c. Overall sexual function	-0.30	A+ R+	P1+ P2+	+ / -
d. IIEF erectile dysfunction	OR 2.5, p=0.0002	A+ R+	P1+ P2+	Y
e. SR pregnancy / miscarriage / stillbirth	OR 1.0, p=0.74	-	-	-

Table 17.12 : Agreement between sexual function measures

	IIEF N=206	Loss of interest in sex N=196	Problems with sexual functioning N=165
IIEF		K=0.46 N=133	K=0.63 N=140
Loss of interest in sex			K=0.54 N=125

Mental health

The results of Chapter 15 (Mental Health) are summarised in Table 17.13.

Table 17.13 : Summary of mental health outcomes

Mental health outcomes	Overall comparison	Group comparison	Subgroup comparison	Dose-resp
a. Kessler 10-item scale	OR 2.6, p<0.0001	A+ R+	P1+ P2+	+ / -
b. General Health Questionnaire 12-item scale	OR 2.4, p<0.0001	A+ R+	P1+ P2+	Y
c. Self-reported physician diagnosis – depression	OR 2.2, P<0.0001	A+ R+	P1+ P2+	Y
d. Self-reported physician diagnosis – anxiety	OR 1.9, p<0.0001	A+ R+	P1+ P2+	+ / -
e. CIDI depressive disorder	OR 1.8, p=0.0005	A+ R+	P1+ P2+	Y
f. CIDI anxiety disorder	OR 1.8, p=0.0001	A+ R+	P1+ P2+	+ / -
g. Neurasthenia	OR 1.9, p=0.01	A– R+	P1+ P2–	N
h. Anti-depressant medication	OR 1.7, p=0.02	A– R+	P1+ P2–	+ / -

The strongest finding is for depression: there is an approximately two-fold increase in depression that is consistent whether depression is measured by a previous physician

diagnosis, by CIDI at the physical examination, or by anti-depressant medications. The agreement across these three measures is shown in Table 17.14. The CIDI depression scale has the greatest sensitivity, picking up the most subjects with depressive disorder; and approximately 70% of those who report a previous physician diagnosis of depression or report taking anti-depressant medications are picked up on the CIDI. Hence the CIDI was taken as the best measure of depression.

Table 17.14 : Agreement between depression measures

	CIDI Depressive disorder N=181	**Anti-depressant medication N=61**	**Physician diagnosed depression N=166**
CIDI Depressive disorder		K=0.25 N=42	K=0.60 N=119
Antidepressant medication			K=0.40 N=55

Two measures of anxiety, previous physician diagnosis and CIDI score, had fair agreement, with a kappa of 0.42. Again, the CIDI scale was more inclusive, detecting more subjects with anxiety than did physician diagnosis, so the former was used as the best measure of anxiety. Both the Kessler and the GHQ are indicators to some degree of depression and anxiety. The kappa for these was fairly good at 0.50, with almost 90% of those who scored poorly on the Kessler also scoring poorly on the GHQ. Since the GHQ detected more subjects as high risk, this was used as the better measure of the two for further comparison. Within the overall mental health domain, we were left with four measures: CIDI anxiety, CIDI depression, GHQ, and neurasthenia. Table 17.15 shows the agreement between these measures.

Table 17.15 : Agreement between mental health measures

	GHQ N=401	CIDI Depressive disorder N=181	CIDI Anxiety disorder N=156	Neurasthenia N=63
GHQ		K=0.24 N=158	K=0.22 N=139	K=0.10 N=59
CIDI Depressive disorder			K=0.39 N=94	K=0.25 N=45
CIDI Anxiety disorder				K=0.23 N=38

Virtually all those who scored high on neurasthenia also scored high on the GHQ (>90%), so these results are subsumed to some degree within the GHQ. There is little overlap between the GHQ, CIDI depression, and CIDI anxiety, so these are preserved separately for further comparisons across domains.

Neuropsychological outcomes

The results of the neuropsychological outcomes (Chapter 16) are summarised in Table 17.16. All four tests for executive function show positive results, although the WAIS Similarities Test is the weakest. The COWAT Letter and Animal tasks show differences against both Amberley and Richmond comparisons and show a dose-response effect, but are not significant in the Program 1 and 2 subgroups. The Trail Making Test Part B shows a difference only against Richmond, but this is consistent in both Program 1 and 2 subgroups, and there is a suggestion of a dose-response effect. The correlations between these four tests are shown in Table 17.17. The correlation coefficients are all weak, indicating that these four tests are targeting different components within the executive function sub-domain. In view of this, the results for the four tests were converted to Z-scores, and the worst score on any of the four tests was taken as representative of this sub-domain.

The three tests for psychomotor speed were also consistently positive, although the inference was weaker than for executive functioning. The results were significant against only one comparison group, the Program 1 and 2 subgroups were not consistent, and the dose-response effect was, at best, equivocal. Table 17.18 shows the correlations for these three tests. These showed fairly good correlations, and again the results were converted to Z-scores, and the worst score on any of the three tests was taken as representative of this sub-domain.

Table 17.16 : Summary of neuropsychological outcomes

Neuropsychological Outcomes	Overall comparison	Group comparison	Subgroup comparison	Dose-resp
Executive Functioning				
a. WAIS Similarities	–0.10, p=0.005	A– R+	P1– P2–	N
b. COWAT Letter Task	–1.77 diff, p=0.02	A+ R+	P1– P2–	Y
c. COWAT Animal Task	–1.12 diff, p=0.0003	A+ R+	P1– P2+	Y
d. Trail Making Test Part B	0.08 diff, p=0.0003	A– R+	P1+ P2+	+ / -
Psychomotor Speed				
e. WAIS Digit Symbol Coding	–3.3 diff, p=0.0005	A– R+	P1+ P2+	+ / -
f. Trail Making Test Part A	0.05 diff, p=0.0015	A– R+	P1– P2+	N
g. Purdue Pegboard Test	–0.69 diff, p=0.049	A– R+	P1+ P2–	N
Attention/Working Memory				
h. Digit Span Forwards Test	-0.20 diff, p=0.40	-	-	-
i. Digit Span Backwards Test	-0.20 diff, p=0.22	-	-	-
New Learning / Memory				
j. WAIS Incidental Learning Free Recall	OR 1.0, p=0.21	-	-	-
k. WAIS Incidental Learning Pairing	OR 1.2, p=0.09	-	-	-
l. WMS Visual Reproduction Immediate Recall	OR 1.05, p=0.05	-	-	-
m. WMS Visual Reproduction Delayed Recall	OR 1.01, p=0.77	-	-	-
n. AVLT Immediate Recall	–0.52 diff, p=0.005	A– R+	P1+ P2–	N
o. AVLT Delayed Recall	–0.51 diff, p=0.03	A– R+	P1– P2–	N
p. AVLT Total Learning	–1.4 diff, p=0.03	A– R+	P1+ P2–	N
q. AVLT Forgetting	0.01 diff, p=0.06	-	-	-
r. AVLT Retrieval Efficiency	–0.04 diff, p=0.08	-	-	-
Visuospatial Ability				
s. Block Design Test	p<0.0001	A+ R+	P1+ P2+	N
Subjective Memory				
t. Memory Complaint Questionnaire (MAC-Q)	4.15 diff, p<0.0001	A+ R+	P1+ P2+	Y

Table 17.17 : Correlations for executive function tests

	COWAT Animal	COWAT Letter	Similarities
Trail B	-0.23	-0.23	-0.18
COWAT Animal		0.37	0.17
COWAT Letter			0.12

Table 17.18 : Correlations for psychomotor speed tests

	Pegboard	Trail A
Digit symbol	0.50	-0.50
Pegboard		-0.41

Three of the nine tests within the new learning/memory sub-domain were positive, although again the inference was weak, with results being significant only against the Richmond group, only in some subgroups, and with no dose-response effect evident. Correlations between these three tests were very strong (see Table 17.19), and again the scores were standardised and the worst score on any of the three tests was taken as representative of this sub-domain.

Table 17.19 : Correlations for new learning/memory

	AVLT Delayed Recall	AVLT Total learning
AVLT Immediate recall	0.84	0.75
AVLT Delayed Recall		0.76

Given that there was no detectable effect for Attention/Working Memory, this left five sub-domains to correlate: executive function, psychomotor speed, new learning/memory, visuospatial abilities (represented by the Block Design test only), and subjective memory (represented by the MAC-Q). The strongest result of all these was the MAC-Q, a validated questionnaire assessing memory complaints. The correlations between these five sub-domains are shown in Table 17.20.

Table 17.20 : Correlations for neuropsychological outcomes

	Psychomotor speed low	AVLT new learning memory low	Visuospatial abilities	MAC-Q (high)
Executive function low	0.36	0.34	0.29	-0.16
Psychomotor speed low		0.28	0.36	-0.24
AVLT new learning memory low			0.24	-0.18
Visuospatial abilities				-0.11

The correlation coefficients are weak, indicating that these sub-domains are independent, and that scoring poorly on one does not mean scoring poorly on another. This again argues for some specificity in that there is not simply a general, non-specific decrease in all cognitive functions; subjects seem to have selective deficits with particular tasks. On the other hand one might argue that a common exposure should affect all subjects similarly, although with something as complex as cognition it would be possible, indeed likely, that education, pre-morbid functioning, etc. would influence how the deficit manifests itself.

The weak correlations across the five sub-domains suggest that these results should be preserved separately for individual comparison **across** domains. However, comparing these five measures across other domains (e.g. SF-36, dermatitis, moles and naevi, sexual function, GHQ, CIDI depression, and CIDI anxiety) indicates that they all have the same pattern of segregation (see Appendix 17A). The magnitude of the correlation was similar for executive function, psychomotor speed, new learning/memory and visuospatial ability (i.e. correlation coefficient ~0.1), but was higher for the MAC-Q (~0.35). In other words, although exposed subjects have different patterns of neuropsychological deficits, they all appear to "cluster" in the same way across the other domains. This means that for the purposes of describing how the deficits "cluster" together, the four sub-domains within neuropsychological outcomes, apart from the MAC-Q, can be collapsed. This was done by taking the worst score on any of these four sub-domains as being representative of the neuropsychological domain, and preserving the MAC-Q score independently.

17.4.2 Across-domain comparisons

Following the within-domain associations, there were 11 measures left to compare across domains: SF-36 physical component score, SF-36 mental component score, physician-diagnosed obstructive lung disease, physician-diagnosed dermatitis, moles/naevi, sexual function, General Health Questionnaire (GHQ), CIDI depression, CIDI anxiety, neuropsychological deficits, and MAC-Q. The correlations across the continuous measures are shown in Table 17.21, and agreement between the dichotomous measures in Table 17.22.

Table 17.21 : Correlations across domains for continuous scores

	SF-36 Mental Component	Neuro-psychological domain summary	MAC-Q
SF-36 physical component	.00064	0.12	-0.35
SF-36 Mental Component		0.17	-0.44
Neuro-psychological domain summary			-0.21

As can be seen from Table 17.21, there is little overlap between the physical and mental component scores of the SF-36 and the neuropsychological tests. There is a slightly better correlation between these three measures and the MAC-Q (subjective memory complaints), although these are still weak to fair. This indicates that these four measures tend to be independent and do not “cluster together” to any significant degree. Table 17.22 also indicates that there is little agreement across domains. CIDI depression and anxiety are somewhat related and both are weakly associated with the GHQ, but there is little agreement across dermatological or respiratory outcomes or across sexual function.

There does appear to be, however, some association between the continuous and dichotomous measures (Table 17.23). Across all the dichotomous domains, those with a positive finding tend to score more poorly on the SF-36 mental and physical component scores (MCS and PCS respectively), neuropsychological tests, and MAC-Q. In most cases,

the means are significantly lower on all the continuous tests for those who have obstructive lung disease, dermatitis, moles/naevi, erectile dysfunction, depression, or anxiety; this is after adjusting for age, rank and posting period.

This association between the dichotomous and continuous measures is supported by a dose-response effect. As shown in Table 17.24, the more dichotomous findings that are present, the poorer the scores on the continuous outcomes. The scores on three of the four scales show a graded and stepwise increase as the number of positive findings increase; the neuropsychological scale shows some irregularity in the graded increase, but the overall trend is similar. The graded increase in scores with increasing number of findings is also similar for the individual sub-domains of executive function, psychomotor speed, new learning/memory, and visuospatial ability (see Appendix 17B).

Table 17.22 : Agreement across domains for dichotomous measures

	Bronchitis / Emphysema N=119	Dermatitis N=183	Other skin lesion N=190	Erectile function N=188	GHQ N=401	CIDI depression N=181	CIDI anxiety N=156
Bronchitis / Emphysema		K=0.12 N=51	K=0.04 N=41	K=0.09 N=47	K=0.02 N=80	K=0.17 N=52	K=0.09 N=38
Dermatitis			K=-0.02 N=55	K=0.18 N=77	K=0.05 N=128	K=0.20 N=76	K=0.11 N=57
Other skin lesion				K=0.04 N=62	K=0.06 N=134	K=-0.01 N=54	K=0.01 N=49
Erectile function					K=0.25 N=152	K=0.23 N=75	K=0.26 N=69
GHQ						K=0.24 N=158	K=0.22 N=139
CIDI depression							K=0.38 N=94

Table 17.23 : Associations across domains for dichotomous and continuous outcomes

Disorder		N	PCS A	p-value	MCS A	p-value	Neuro low	p-value	MAC-Q tot	p-value
Bronchitis/ Emphysema	**Yes**	119	39.93	0.13	37.96	<0.0001	-1.53	0.24	29.02	0.001
	No	468	42.25		44.31		-1.39		26.97	
Dermatitis	**Yes**	183	39.51	0.002	38.98	<0.0001	-1.55	0.04	29.02	<0.0001
	No	401	42.79		44.95		-1.36		26.66	
Other skin lesion	**Yes**	190	40.50	0.07	41.14	0.04	-1.54	0.26	27.93	0.39
	No	422	42.33		44.14		-1.37		27.12	
Erectile function	**Yes**	188	38.41	<0.0001	36.62	<0.0001	-1.66	0.006	29.55	<0.0001
	No	308	43.58		47.87		-1.26		25.97	
GHQ	**Low (better)**	207	43.76	0.001	53.50	<0.0001	-1.28	0.01	24.87	<0.0001
	High	401	40.73		37.72		-1.49		28.69	
CIDI Depression	**Yes**	181	39.73	0.009	33.23	<0.0001	-1.67	0.002	29.75	<0.0001
	No	416	42.66		47.39		-1.33		26.39	
CIDI Anxiety	**Yes**	156	38.25	<0.0001	32.20	<0.0001	-1.66	0.001	30.56	<0.0001
	No	441	43.00		46.91		-1.35		26.32	

Table 17.24 : Dose-response effect, relating the number of positive outcomes to scores on the continuous scales

All outcomes	N	PCS A	MCS A	Neuro low	MAC-Q tot
0	84	44.34	55.33	-1.18	24.21
1	133	44.64	51.32	-1.25	24.32
2	129	43.19	44.62	-1.46	27.97
3	128	40.51	40.88	-1.34	28.04
4	88	38.91	30.73	-1.72	30.90
5	35	35.59	28.07	-1.64	31.15
6	17	34.90	26.37	-2.55	31.63
7	2	23.93	24.18	-1.46	35.00

17.5 Discussion

17.5.1 Caveats

Before discussing the overall significance of these results it is important to highlight some of the study's limitations. Attempting to identify a group of people who participated in DSRS activities, sometimes up to 30 years after the fact, when no consistent records were kept, is methodologically very difficult. It is impossible to test the completeness of the DSRS personnel list.

It is also important, although obvious, to acknowledge that the current study does not include those who have died since participating in DSRS activities. The Mortality and Cancer Incidence Study conducted for Phase 2 of SHOAMP identified a substantial survivor bias. This suggests that those surviving for inclusion in the current health study may skew the results towards the null, i.e. decrease the chances of finding a significant result.

It is also important to acknowledge potential participation bias. Participation rates were 40%, 48%, and 77% for the Amberley, Richmond, and DSRS groups respectively, despite multiple methods of follow-up. Those in the comparison groups who participated in the study were older compared to those who did not, and they were involved in earlier DSRS programs. Of those who did participate, the comparison groups were also older on average than the DSRS group. All these biases would tend toward the null, decreasing the chances of finding a significant result between groups.

On the other hand, it is possible that those who were sicker tended to participate more willingly in the study. If these were enriched in the DSRS group, it would tend to bias away from the null; if these were enriched in the comparison groups, it would tend towards the null. It is difficult to ascertain the overall effect of all these biases on the results. Given that the three study groups were reasonably well matched, and that all analyses were adjusted, at a minimum, for age, rank, and posting dates, as well as other pertinent potential confounders, we believe that we have taken the best steps available to us to minimise the potential for results to be affected by such biases. It is worth mentioning that although the current study enrolled only those who participated in a formal DSRS program, the results may be relevant to those personnel who were involved in *ad hoc* or "pick and patch" type repairs.

17.5.2 Relative associations

It is worth noting that the frequency of every self-reported symptom was increased in the DSRS group relative to both comparison groups. In some cases, these symptoms were not supported by objective test results. They included:

- neuropathic symptoms and biothesiometry
- self-reported wheezing and spirometry.

In other cases, these symptoms *were* corroborated by objective or validated measures. They included:

- rash/dermatitis and physical examination
- sexual dysfunction and IIEF scores
- depression and anxiety, and CIDI scores
- memory/cognitive problems and neuropsychological tests.

In all self-reported symptoms, the association was strong ($OR \geq 2$), significant against both comparison groups, consistent between Programs 1 and 2, and showed some dose-response effect, making these findings the ones with the strongest inference in the entire study.

Due to the self-report nature of symptoms, it is impossible to rule out the possibility of some over-reporting of symptoms in the DSRS group and/or perhaps some under-reporting in the comparison groups. While self-reported symptoms are subjective, the validity of these symptoms is nevertheless bolstered by the fact that participants did not exhibit a non-specific, general increase in all symptoms. Some symptoms – e.g. loss of concentration, difficulty finding the right word, and forgetfulness – were quite specifically increased. In addition, there was little association between the self-reported symptoms – e.g. in the cardiovascular domain, subjects did not typically report all symptoms as present. This argues for some specificity of the association between exposure and symptoms.

17.5.3 Strong associations

The strongest findings in the study relate to quality of life, sexual function, mood, and memory. The physical and mental components of the SF-36 were significantly decreased in the DSRS group, placing them in the lowest 20^{th} –30^{th} percentile of the Australian population. Depression and anxiety were increased almost two-fold; and sexual dysfunction was

increased 2.5-fold versus comparisons, even after controlling for mood disturbances. Memory impairment, as assessed by a validated questionnaire (MAC-Q), was increased four-fold compared to the Amberley and Richmond groups. These findings were consistent in that, although those who were depressed differed from those who were anxious and from those who had sexual dysfunction, all of these groups scored worse on physical and mental quality of life, and worse on subjective memory. All three groups also scored worse on neuropsychological outcomes.

17.5.4 Moderate associations

A number of findings were judged to be of moderate strength. These included dermatitis, moles/naevi, chronic obstructive lung disease (bronchitis/emphysema), and tests within the neuropsychological domain. There was a 1.5 to 2.5-fold increase in dermatitis, with a positive correlation, albeit weak, between the self-reported rash, previous physician-diagnosed dermatitis, and dermatitis observed at the physical exam. There was also an approximate two-fold increase in moles/naevi and obstructive lung disease, although these did not overlap with each other or with dermatitis. All of these, however, did score worse on neuropsychological function. The strongest association in this domain was for executive function, with all four tests showing a decreased performance of the DSRS group in this sub-domain; this decrease in performance was equivalent to a one decile drop in normative values. The psychomotor speed tests (n=3), the new learning/memory tests (n=3) and the visuospatial test results were of weaker inference, in that they tended to show significance against only one comparison group, to be significant in only one sub-group comparison, and to have, at best, an equivocal dose-response effect. There was no overlap in these four sub-domains of neuropsychological function, indicating that subjects had differing patterns of cognitive loss. The impact on new learning/memory is perhaps the weakest in this group, in that the other five tests (out of eight) in this sub-domain did not detect any association. On the other hand, the results are bolstered by the fact that adjustment for depression and anxiety did not reduce this association.

17.5.5 Weak associations

Colour vision deficiency had a weak association with exposure. The colour vision tests gave a positive result only when analysed using the colour confusion index, not by clinical diagnosis. This was only significant against Richmond but not in subgroup comparisons, and there was an equivocal dose-response effect.

17.5.6 No evidence of association

A number of measures showed no evidence of association, including blood pressure drop, asthma, vibration sense in the extremities, olfaction, and attention/working memory, as well as standard haematological evaluation.

17.5.7 Associations across domains

There is little overlap in the positive findings such as dermatitis, moles/naevi, obstructive lung disease, colour vision, depression, anxiety, or sexual dysfunction. However, those with at least some positive finding in these areas consistently tend to have a poorer quality of life (physical or mental), poorer memory, or poorer neuropsychological function, indicating that there is some consistency to the data.

17.5.8 The evidence *in toto*

The strongest associations in the current study relate to the nervous system, i.e. mood, cognition, and memory, as well as quality of life, and erectile function. Gobba[2,3] has suggested that colour vision and olfaction might be used as sensitive indicators of exposure to neurotoxins. The lack of strong effect in these domains might therefore be seen as weakening the case for an F-111 DSRS association. In addition, the lack of adjustment for multiple comparisons raises the possibility that some of these positive results may be spurious. Furthermore, the lack of a cohesive set of findings, i.e. a “typical picture”, and the paucity of positive results on “hard/objective” measures, may be seen as weakening the likelihood of an effect.

On the other hand, the lack of a “typical picture” may not be surprising given the complexity of the construct being measured. We postulate that with so many different sub-domains within neuropsychological function, and so many different potential exposures and routes of exposure, each person may manifest neurotoxic effects in slightly different ways depending on education, personality, pre-morbid function, etc.

With regard to the possibility of chance positive findings, the positive results tend to be corroborated through various independent tests, e.g. four tests of executive function all show

the same association. The data also show internal consistency in that: a) associations with quality of life, neuropsychological outcomes, or memory complaints across various domains are similar; and b) as the number of positive findings increases, the mean scores on quality of life, neuropsychological tests, and subjective memory also worsen in a graded, dose-response effect.

The cognitive and sexual function outcomes also remain positive after adjustment for a number of potential confounders, including depression and anxiety as measured by the CIDI score. This seems to indicate that the neuropsychological outcomes and erectile problems are not simply a by-product of the mood disorder, nor mediated simply through the mood disorder, but that there may be a direct organic/neurotoxic effect. Alternatively, it is also possible that these findings may be due to a behavioural response to a difficult situation, i.e. hot, cramped environment, offensive-smelling chemicals, etc. Such a mechanism was postulated for the “aerospace syndrome”, in which a group of workers in an aerospace factory working with epoxy resins reported a number of cognitive and mood abnormalities.[4] Regardless of the mechanism of effect, the significance remains the same. Although some might be tempted to discount the import of the main findings as relatively “soft” outcomes, and not perhaps as convincing as an objectively-measured change in vibration sense for example, the main outcomes are all measured with widely-used, validated scales that are accepted and recognised. For instance, the SF-36 quality of life scale has been shown to be a significant predictor of future health risk and morbidity.[5-7]

An additional point of note is that most differences were significant against both comparison groups or against the Richmond group only. The Richmond group had similar technical trades and aircraft maintenance work as the F-111 DSRS group, but no F-111 DSRS involvement, and did not show similar results to the exposed group. This argues for some specificity of the association with DSRS activities, and supports (but does not prove) the hypothesis that the “causative” agent(s) responsible for ill-health effects were particular to the F-111 program and were not general to aircraft maintenance activities. In addition, the strongest effects are consistent across Program 1 (which also includes many Wing Program participants) and Program 2 (which includes many Spray Seal Program participants), which suggests that the “causative” agent(s) were common to at least these two programs.

Although the aim of the current study was to ascertain whether those involved in DSRS activities had adverse health outcomes, there were some *a priori* hypotheses about the possible ill-health effects of some exposures. The internal investigation and the Board of Inquiry both raised concerns about solvents and isocyanates. Some expected long-term

effects of solvents were seen, for example dermatitis and neuropsychological changes, but other effects, for example hepatic or renal changes, were not. The finding that some solvent-compatible effects were significant against the Richmond but not the Amberley comparison group is unexpected but may be due to participation bias and residual confounding.

A priori, a number of other exposures were of concern, some of which were unique to particular programs. However, the pattern of results showed no demonstrable differences between the participants in programs 1 and 2.

There are additional analyses that may shed further light on causation but which were beyond the scope of this study. The associations, consistency, and patterns within the results may be more fully explored with a full factor analysis, to identify the “axes” on which the findings segregate. Moreover, although the within-domain and across-domain correlations were carried out only within the DSRS group, it is possible to perform the same correlations in the two comparison groups separately. This would give additional evidence regarding causation. The logic behind this is as follows: if DSRS exposure were causative for symptoms in two separate domains, e.g. dermatitis and poor quality of life, then these two findings would correlate highly. If dermatitis and poor quality of life in the comparison groups were not caused by DSRS (and indeed they should not, given that they had no exposure) then these two findings need not be linked and should not necessarily correlate. Hence the pattern of associations between the exposed and comparison groups should provide further evidence regarding causation.

17.6 Conclusions

There are unavoidable uncertainties in the interpretation of the study results, due to such factors as uncertain sampling frames, potential survivor bias, low participation rates, and multiple comparisons. Nonetheless, the results point to an association between F-111 DSRS involvement and poor physical and mental quality of life, erectile dysfunction, depression, anxiety, and subjective memory impairment.

17.7 References

1. Fleiss, J.L. Statistical methods for rates and proportions. in *Statistical Methods for Rates and Proportions.* (ed. Fleiss, J.L.) (John Wiley and Sons., 1981).
2. Gobba, F. & Cavalleri, A. Color vision impairment in workers exposed to neurotoxic chemicals. *Neurotoxicology* **24**, 693-702 (2003).
3. Gobba, F. Colour vision: a sensitive indicator of exposure to neurotoxins. *Neurotoxicology* **21**, 857-862 (2000).
4. Sparks, P.J. et al. An outbreak of illness among aerospace workers. *West J Med* **153**, 28-33 (1990).
5. Sprenkle, M.D., Niewoehner, D.E., Nelson, D.B. & Nichol, K.L. The Veterans Short Form 36 Questionnaire Is Predictive of Mortality and Health-Care Utilization in a Population of Veterans With a Self-Reported Diagnosis of Asthma or COPD. *Chest* **126**, 81-89 (2004).
6. Fan, V.S. et al. Using quality of life to predict hospitalization and mortality in patients with obstructive lung diseases. *Chest* **122**, 429-436 (2002).
7. Elliott, T.E., Renier, C.M. & Palcher, J.A. Chronic pain, depression, and quality of life: correlations and predictive value of the SF-36. *Pain Medicine* **4**, 331-339 (2003).