

Contents - Chapter 2

Methods

- 2.1 Overview of neurocognitive tests. p. 63**
- 2.1.1 Tests of short term memory.
- 2.1.2 Tests of information processing.
- 2.1.3 Tests of general intellectual skill.
- 2.1.4 Tests of psychomotor speed.
- 2.1.5 Tests of visuo-spatial perception and construction.
- 2.1.6 Tests of working memory.
- 2.1.7 Tests of supraspan learning.
- 2.1.8 Beck Depression Inventory.

- 2.2 Overview of objective and subjective tests of somnolence. p. 73**
- 2.2.1 Epworth Sleepiness Scale.
- 2.2.2 Steer Clear Driving Simulator.

- 2.3 Enrollment criteria. p. 74**
- 2.3.1 Neurocognitive dataset.
- 2.3.2. Melatonin dataset.

2.4	Polysomnographic data for the neurocognitive dataset.	p. 76
2.4.1	Collection of data.	
2.4.2	Analysis and scoring of respiratory events, sleep stages and arousals.	
2.4.2	Subanalysis of arousals. A modified scoring system.	
2.4.2.1	Rationale for development of a four level scoring system.	
2.4.2.2	Method of scoring arousals by modified criteria.	
2.4.2.3	Validation procedure for modified arousal scoring.	
2.4.3	Analysis of measures of hypoxia.	
2.5	Neurocognitive battery administration.	p. 81
2.6	Statistical methods for analysis of neurocognitive outcomes.	p. 83
2.6.1	Simple regression modelling.	
2.6.2	Multiple and stepwise regression techniques.	
2.6.3	Secondary analysis of data. Determination of risk ratios.	
2.6.4	Validation procedure for the modified arousal scoring method.	

2.7	Melatonin samples.	p. 86
2.7.1	Collection methods.	
2.7.2	Analysis methods.	
2.7.2.1	Urinary 6-sulphatoxymelatonin analysis.	
2.7.2.2	Polysomnographic analysis of hypoxia and arousal.	
2.7.3	Statistical analysis of data.	

METHODS CHAPTER

Introduction.

The data with respect to neurocognitive dysfunction in OSA was collected from 55 patients over 18 months (June 1995 - November 1996). All melatonin samples were collected over the periods April-July 1994-1996 from a separate group of 40 subjects.

The following chapters will categorise the neurocognitive battery, the methods of administration, scoring and statistical analysis. Similar attention will be paid to the polysomnographic analyses. Furthermore, additional chapters will elucidate the melatonin sample collection methods and data analysis.

2.1 Overview of Cognitive Tests.

For the sake of simplicity the test battery order is presented in List 1 (see Chapter 2.5). All tests were performed on the morning immediately following the sleep study. The test time ran from 0800 to 1100 hours. Further detail is presented in Chapter 2.5. In order to allow for the development of a clinical relationship between cognitive tests and the measured components of sleep disordered breathing, the tests have been compartmentalised into broad categories. The specifics of the tests within each category are discussed separately.

2.1.1 Tests Of Short Term Memory.

- Logical memory passages I and II
- Benton Visual Retention Test
- Digitspan forwards

The logical memory passages of the Wechsler Memory Scale - Revised (WMS-R)¹³⁶ are two brief stories incorporating details such as names, places and numbers interspersed with contextual text. Each story is read in turn and the subject is required to recall as much detail as possible. Marks are accorded for correctness and accuracy allowing for a minor degree of interpretation. The standard scoring procedure and normative data have been well validated and published (Wechsler 1987). A score out of a possible best of 25 correct responses can be calculated and percentiles derived from age-matched controls. The candidate is then instructed to remember both story passages and delayed recall is then tested at 30 minutes. There is no re-reading of the passages prior to the delayed recall testing. Scoring of the re-testing procedure is as previous. All tests were performed with modified Australian content as described by Sullivan¹³⁷.

The logical memory passages were chosen for ease of administration and have been demonstrated to show good internal validity¹³⁸.

The Benton Visual Retention Test (BVRT) is a complex test of visual recall and reconstruction. Described by the author¹³⁹, it is both complex in administration and scoring. Normative data¹⁴⁰ on 1,128 healthy normals has been validated demonstrating a positive association between education level and performance. The test comprises the presentation of geometric objects, squares, circles etc of increasing complexity. A 10 second presentation is followed by a period

whereby the subject is requested to reconstruct the picture they have been shown. The total score is derived from a complex scoring of the correct and erroneous reconstructions. The final outcome is measured by a calculation of the predicted score for age and level of education and the measured score. The test has been previously used in a study of snorers versus normals⁷⁷.

Digitspan forward¹⁴¹ of the WMS-R is a test of immediate numerical recall. The subject is given a series of numbers of increasing length (3 - 8 digits). They are required to retain the sequence and repeat it in exactly the same order. Two different sequences are presented at each length. The score is derived by the total correct (out of 12) and the test terminated once an incorrect sequence has been given at both presentations of a given digit length. Normative scoring for age is then calculated¹⁴¹.

The tests outlined above were selected in an effort to measure short term memory by three different presentation modes, numerical, visual and verbal, since selective memory dysfunction may occur in a given clinical setting.

2.1.2 Tests Of Information Processing

- Symbol Digit (oral presentation)

The Symbol Digit Test¹⁴² is a useful measure of cognitive processing speed. The subject is presented with a key of numbers (1 - 9) matched to symbols of a simple design. Below the key is a series of matching boxes (15 per line) with symbols presented above empty boxes. The oral performance of the test follows an untimed attempt to match the first 10 symbols by calling out the numbers from the key above. No psychomotor component is incorporated by this

technique. Following the trial, the subject is given 90 seconds to call out the numbers in sequence, corresponding to the symbols in each box. A score of all correct responses is then normed for age¹⁴².

The test as a measure of information processing speed has been validated both in normals¹⁴³ and patients with closed-head injury¹⁴⁴.

2.1.3 Tests Of General Intellectual Ability / IQ.

- Vocabulary component of the Shipley Institute of Living Scale (SILS)
- Abstraction component of SILS
- Estimated WAIS full scale IQ from SILS

The vocabulary component of the SILS is an untimed word matching test. The subject is given a key word and asked to find the appropriate synonym from four given choices. Forty words are presented and a score calculated from the total of correct answers. A t-score is then derived from the accompanying text manual¹⁴⁵ as corrected for age. Vocabulary as a fixed measure of intellectual ability is generally recognised to be unimpaired by cognitive insult, thus the derived Vt was used as a regression covariable to control for general intellectual ability across all other cognitive measures.

The abstraction component of the SILS is a 10 minute timed test of problem solving. The subject is given a series of words, letters or numbers followed by blank spaces. They are instructed to examine the pattern of the preceding series in order to complete the solution. A total of 20 puzzles are given with an instruction to attempt all problems in the time allocated. The total of correct answers is used to derive a t-score from normal tables for age matched controls.

A further subanalysis is possible using predicted values for the test based on the level of education and the Vt. This is used to derive the abstraction quotient again from the normed tabular data.

The WAIS full-scale IQ is calculated from the sum of the t-scores for vocabulary and abstraction using normal tables published with the manual for the test. The use of this test as a comparable measure of IQ has been validated by Dalton¹⁴⁶ particularly for grouped data.

The controlled oral word association test (COWAT) is predominately a test of verbal knowledge although unlike the SILS vocabulary component, it is a timed test requiring reasonable verbal agility. The subject is presented in turn with three consonants: F, A & S. They are given 60 seconds to call out as many words as possible starting with the given letter. Proper nouns, perseveration and words of like structure (eg eat, eats, eaten) are scored correctly for the initial word only.

A normative value is derived from age-matched totals for the sum of the three trials. The test has been demonstrated by Axelrod et al¹⁴⁷ to be unrelated to intellectual competence, educational experience or general health status of the subject, although there is age related decline. Furthermore verbal knowledge has been demonstrated to be a major contributor¹⁴⁸.

2.1.4 Tests Of Psychomotor Speed.

- Grooved peg board
- Block design test
- Trails making tests parts A and B

The Grooved peg board¹⁴⁹ is a test of psychomotor speed and manual dexterity. The subject is instructed to place grooved metal pegs into randomly rotated matching grooved holes, set out in a 5 by 5 matrix. They are instructed to use the dominant hand first with subsequent non-dominant re-testing. The time to completion is measured and normal values derived from reference tables for each hand¹⁴⁹. A subcalculation for dropped pegs can be included, although the time penalty for continuing the test is deemed equally satisfactory.

The block design test is a complex measure of psychomotor speed as well as visuospatial perception and construction. The subject is given a set of 4 red/white blocks identical in design. They are shown 5 consecutive patterns (2-dimensional) and are directed to recreate the pattern using the coloured blocks (3 dimensional). The time to completion is recorded, and a score for each trial is derived from 3 time intervals - higher marks accorded for faster performance. The same paradigm is then performed using 9 blocks over 4 trials, with longer time intervals being allowed for each test.

The total score for all tests is then calculated and normative data corrected for age is derived from known tables¹⁵⁰. Indeed, while image segmentation processes remain constant with age, other processes become slower¹⁵¹ such that age corrections become critical to analysis of data derived by this test.

Trails making tests part A and B are timed tests of visual tracking and sequencing.

Part A: The subject is instructed to join numbered circles in ascending order e.g. 1 to 2, 2 to 3 , 3 to 4 etc. The circles are spread across the page in random order, but in such a manner that no lines are ever crossed. Time from start to completion is recorded and normal values are derived from matched controls for age and level of education.

Part B: Follows a similar paradigm to the first part to the test but incorporates an added task of mental dexterity. The subject is required to join the circles in sequence, however they must simultaneously progress through a concurrent alphabetic process i.e. 1 to A, A to 2, 2 to B, B to 3 etc. Normal values are derived for age and level of education from tabular data¹⁴².

Trails B is deemed harder¹⁵² since it demands greater levels of motor speed and visual search. Furthermore it is more closely associated with visual non - verbal intelligence than with attention information processing. For subjects of average or higher intelligence, test performance is independent of intellectual ability¹⁵³ and demonstrates good inter-rater variability¹⁵⁴.

2.1.5 Tests Of Visuospatial Perception and Construction.

- Reye - Osterrieth Complex Figure Test

The Reye-Osterrieth Complex Figure Test (RCFT) is an untimed test of visuospatial perception and construction¹⁵⁰.

The subject is shown a complex line diagram comprising several standard geometric patterns including squares, circles, oblongs and triangles. Within the construct are smaller additional details such as parallel lines and dots. There are, in total, 18 standard elements. The subject is asked to reproduce the figure as neatly and accurately as possible in a space on the paper below the template. No rubbers or rulers are allowed. A complex scoring algorithm is employed with marks out of a total of 36 being accorded. Deductions are made for incorrect placement, missing detail and poor spatial perspective. The scoring method has

been validated by Loring¹⁵⁵ who demonstrated that the criteria could be reliably applied by independent raters.

2.1.6 Tests Of Working Memory.

- Paced Auditory Serial Addition Task (PASAT 2.4 second presentation)
- Digitspan - reverse.

The PASAT¹⁵⁰ is a timed test of mental addition. The subject is presented with sequential single digit numbers read from an audio cassette at 2.4 second intervals. They are instructed to add the last two digits that have been presented. After each digit is presented they are asked to call out the answer from the calculation performed. Thus, there is both a timed addition task as well as a short term memory task for the retention of the previous number. The test has been demonstrated by Deary et al¹⁵⁶ to correlate with all WAIS-R subtests. No effect of education or race has been demonstrated^{157, 158} although age and IQ have been shown to affect the PASAT test results and should therefore be controlled for in analysis.

The reverse presentation of digitspan¹⁴¹ (description Chapter 2.1.1) is a test of numerical short term memory and attention. The subject is presented with a series of single digit numbers of progressively increasing length, ranging from 2 - 7 digits. They are required to remember the numbers as presented and call them back to the instructor in exactly the opposite order. For each digit length they are given two trials. The test is terminated when an incorrect answer is given for both trials of a given length. A total out of 12 is calculated and normal values derived from age-matched tables¹⁴¹.

The test has been demonstrated to be age sensitive in patients over 60 years, possibly due to a decreased flexibility in processing changes¹⁵⁹. This needs to be considered when using the test in older patients.

2.1.7 Tests Of Supraspan Learning.

- Serial digit learning

Serial digit learning is a test of verbal memory which incorporates an element of learning strategy differentiating it from digitspan.

The subject is presented with a very long number either 8 or 9 digits depending on age and level of education. Subjects over 65 or with less than 12 years of schooling receive an 8 digit test, all others receive 9 digits. They are instructed in advance that they will be told a long number of a given length. The number is then read out at a regular speed and without pause. There are no repeated digits. The subject is advised to try and “learn” the first part of the number and add on numbers with subsequent attempts. They are given 12 opportunities to complete the task. The test is terminated when the correct sequence is given by the subject at two consecutive attempts. A maximum total of 24 is calculated for the number of correct answers ie 2 marks for each attempt with 1 mark being accorded for any attempt with only one error and none for any trials with greater than two errors. Normative results are derived from tables of age-education-matched normals¹⁶⁰. The test has been demonstrated to correlate with attention-information processing rather than general memory¹⁶¹ given the requirement for strategy development such as linking the 9 digit sequence into 3 blocks of 3.

2.1.8 Beck Depression Inventory (BDI).

The BDI ¹⁶² is a self administered, untimed questionnaire of mood state. The subject is required to complete a 21 item inventory of subjective rating pertaining to symptoms and attitudes. The test has been designed to assess the severity of depression in adolescents and adults and has been validated for use in detecting possible depression in normal populations¹⁶³. It has been demonstrated to correlate with other standard tests of mood such as POMS (Profile of Mood States)¹⁶⁴. The test, like many self rating mood scales, includes embedded questions regarding sleep, weight and sexual interest which are likely to score abnormally in OSA patients. Unfortunately, more disease specific questionnaires are still undergoing validation and were not available for use in this study. Other, perhaps more sensitive tests, such as the Multiphasic Minnesota Personality Inventory were deemed to be inappropriately lengthy. Generic questionnaires such as Short Form 36 may have application in the setting of OSA but do not specifically reference depression in a measurable manner.

2.2 Overview of Objective and Subjective Tests of Somnolence.

2.2.1 Epworth Sleepiness Scale.

This self reporting 8 point questionnaire was developed as a simple and quick assessment of subjective sleepiness. The subject is asked to rate (from 1 - 3) the likelihood of falling asleep in circumstances of low stimulation (eg reading a book, passenger in a car). A total from 0 - 24 is then calculated, with greater scores correlating with increased somnolence. It has been validated against objective sleepiness scores, in particular the MSLT²⁷. It has also been demonstrated to be an effective tool for measuring persistent daytime somnolence in adults²⁹.

2.2.2 Steer Clear Driving Simulator.

This relatively simple computerised test was initially designed to measure vigilance impairment³⁰.

The subject is presented with a two-dimensional, 2 lane road intended to represent a monotonous highway drive. During a 30 minute study, 780 obstacles (Steers) are pseudo-randomly presented. The subject is instructed to hit the 'space bar' of the key board to shift the car from one lane to the other to avoid the oncoming obstacle.

There is no standardised data with respect to computer processor speed, screen size or speed of presentation of the obstacles. For the work presented here, the test (version B) was run on a 486 NEC computer with 13 inch colour screen at the default setting of 70.

Steer Clear has been demonstrated to correlate with a higher rate of automobile accidents in OSA patients who have been shown by comparison to normals to perform more poorly on this driving simulator.

A total number of 'hits' is calculated by the software and z-scores derived from age-sex-matched normative data¹⁶⁵.

The sensitivity of Steer Clear for predicting automobile accidents remains unknown. Furthermore, no studies to date have sought to correlate the test with subjective somnolence (see results). Despite these flaws, it remains a simple and useful tool for objectively measuring vigilance. It was selected for use in this study given its value in determining vigilance decrement in OSA, as has been demonstrated by Engleman et al¹⁰⁹ who showed improvement post CPAP treatment.

2.3 **Enrollment.**

2.3.1 **Neurocognitive Data Set.**

PATIENTS

55 male volunteers were randomly recruited through the Sleep Unit of the Royal North Shore Hospital. All patients were undergoing polysomnography for investigation of possible sleep disordered breathing. The primary investigators had no prior knowledge of the magnitude of sleep disturbance at the time of administration of the test battery. Consent was obtained prior to test administration and in accordance with the Ethics Committee of the institution. All subjects were asked to refrain from taking caffeinated beverages on the morning of testing.

EXCLUSION CRITERIA

Patients older than 70 years or with a past history of excessive alcohol, epilepsy, previous head injury, or past history of stroke were excluded from the study. Patients were allowed to use their regular medications. Patients taking benzodiazepines or anti-depressants were not enrolled.

2.3.2 Melatonin Data Set.

PATIENTS

40 randomly selected males aged between 18-65 were asked to participate. Patients were enrolled from the Sleep Investigation Unit of the Royal North Shore Hospital. All patients completed a consent form under the guidelines of the Ethics Committee of the Institution.

EXCLUSION CRITERIA

Patients with a history of severe underlying respiratory or cardiac failure, the presence of right or left heart failure on physical examination, documented or clinically evident thyroid or adrenal disease were excluded. Patients on oral steroids or those using tricyclic antidepressant or benzodiazepine medications were excluded so as not to interfere with assay techniques. Female patients were excluded to avoid menstrual hormonal fluctuation which could alter melatonin levels.

2.4 Polysomnographic Data for the Neurocognitive Data Set.

2.4.1 Collection of Data.

All studies were performed using standard 21 channel computerised polysomnography (Compumedics - Melbourne).

Collected data included:

- EEG (Single channel C3, A1/C4,A2)
- Submental EMG
- EOG
- Thoracic and abdominal movement by inductive plethysmography
- Airflow by pressure transducer
- Finger oximetry
- Tibialis EMG

2.4.2 Analysis and scoring of respiratory events, sleep stages and arousals.

Apnoeas and hypopnoeas were defined as a greater than 10 second obstructive apnoea (complete cessation of airflow) or obstructive hypopnoea (50% or greater reduction in tidal volume). All respiratory events required a $\geq 4\%$ desaturation to be included in the analysis. The total apnoeas and hypopnoeas for REM and NREM sleep were derived separately. An average across the night score was calculated to derive the total RDI.

All polysomnographs were scored by one of two full time sleep technologists. Previous (unpublished) inter-rater reliability for the laboratory has been

calculated at > 90%. This process was performed as part of ongoing laboratory quality assurance.

Sleep stages were scored according to R & K criteria, defined as:

- 1: Low voltage mixed frequency with or without slow rolling eye movements.
- 2: Presence of spindles or K-complexes.
The absence of the above excluded epochs from Stage II.
- 3: Presence of $\geq 20\%$ and $<50\%$ delta waves per epoch.
- 4: Presence of $\geq 50\%$ delta waves per epoch.

REM: Low voltage mixed frequency.

Presence of spindle or K complexes in the first half of any epoch was rescored as Stage II.

Reduction in chin EMG signal was required and fast eye movements on the EOG needed to be present.

Arousals were scored according to the Atlas Task Force of the American Sleep Disorders Association¹⁶⁶. In summary, arousals were defined as a ≥ 3 second change in EEG sleep state with increase in chin EMG irrespective of the cause of the arousal.

2.4.2 Subanalysis of Arousals. A Modified Scoring System.

2.4.2.1 Rationale for Development of a Four Level Scoring System.

Previous data by Stepanski et al³⁸ demonstrated that when arousals were subclassified by duration and magnitude, that a strong correlation existed between sleepiness and arousals of shorter duration. Furthermore, there was an additional correlation between sleepiness and the frequency of these brief arousals. This work was later confirmed by Roehrs et al⁴¹ in a sleep fragmentation model. More recently, Martin et al²¹ demonstrated both increased sleepiness and cognitive decrement after one night of sleep fragmentation by auditory tone (using a ≥ 3 second change in EEG to define an arousal).

Based upon the findings of these previous authors, a modified version of Stepanski's criteria was evolved in an effort to define whether the outcomes of this study correlated with arousals of a specific duration or magnitude. It was hoped that the presence of a larger sample size and use of a more expansive neurocognitive battery than the previous studies would generate greater sensitivity and specificity in detecting arousal driven cognitive decrement.

The arousal scoring system was developed by combining the IA and IB, IVA and IVB definitions of Stepanski. This was in accordance with the finding of Stepanski that no extra information was forthcoming with the use of further sub-classification. The follow criteria were applied:

LEVEL	CRITERIA
I	Increase in EEG frequency and increase in chin EMG 3-10 seconds
II	Alpha burst in EEG of 10-29sec.
III	A stage shift to a lesser stage (eg. 2 to 1)
IV	An awakening from any stage of sleep (1-4 and REM)

An awakening was defined as 30sec. of wakefulness in the EEG with/without eyes open.

2.4.2.2 Method of Scoring Arousals by Modified Criteria.

All studies were independently scored by one of two sleep technologists according to the standard criteria of the Atlas Task Force and the R & K criteria. All studies were then copied to optical media and held by the primary investigator. At study termination, all PSG results were relabeled by the sleep technologists with deletion of all names and dates from the original studies and a randomised coding applied. All studies were then rescored by the primary investigator having been blinded to the coding procedure. All original arousals were rescored according to the modified Stepanski criteria and assigned to one of the four levels outlined in Chapter 2.4.2.1.

2.4.2.3 Validation Procedure for Modified Arousal Scoring.

A random series of 8 studies were similarly presented to another physician (neurologist) for rescored according to the modified criteria. Inter-rater variability was determined by simple linear regression and paired t-testing. For details of the statistical methods of validation see Chapter 2.6.4.

Eight studies were rescored by the primary investigator and similar analysis of variance derived for intra-rater variability. Having determined good inter and intra-rater concordance, the total number of each arousal type was then subjected to analysis against cognitive and somnolence outcomes.

2.4.3 Analysis of Measures of Hypoxia.

The following variables were extracted from the computerised polysomnograph:

- Lowest saturation in NREM.
- Lowest saturation in REM.
- Average desaturation for the entire night.
- Total time (minutes) with a saturation below 90%.

All values were derived by the software of the sleep system. The total time with a saturation below 90% was later logarithmically transformed to normalise the data given the spread of values.

Previous authors such as Bedard⁷⁹ defined the magnitude of hypoxia at an arbitrary level below 80%. This cut-off would define more severe OSA and was not considered suitable for this study. Furthermore, a saturation of 90% is at the steepest part of the oxygen saturation curve and standard practice accepts that tissue hypoxia is likely to occur below this physiological threshold. Hence, a saturation cut-off at 90% was deemed to be clinically appropriate.

Findley⁷⁵ determined a correlation existed between the median nocturnal saturation and the cognitive outcomes. As the total area under the curve may be more representative than a specific point of greatest frequency on the curve itself, it was felt that total time may be more reflective of the magnitude of the hypoxic injury.

2.5

Neurocognitive Battery Administration.

The following list defines the sequence of the test battery.

Order and timing of the test battery

Epworth Sleepiness Scale

Steer Clear Driving Test

5 minute break.

Digitspan forward and reverse

Trails A and Trails B

Digit Symbol substitution (oral)

Controlled Oral Word Association Test

Reye-Osterrieth Complex Figure Test

Serial digit learning

Grooved Peg Board (dominant then non-dominant hands)

Beck Depression Inventory

Shipley Institute of Living Scale (verbal untimed, then timed abstraction)

10-15 minute break.

The following tests were performed by Mr. Geoffrey Marshall, neuropsychologist, Royal North Shore Hospital.

Logical Memory Passages I

Benton Visual Retention Test

Block Design

PASAT (Paced Auditory Serial Addition Task - 2.4 second)

Logical Memory Passages II

All patients were requested to refrain from ingestion of caffeinated beverages on the morning of testing. Polysomnography was terminated at 0600 hours and the patients woken to attend to personal hygiene and to allow time for a light breakfast. Testing commenced at 0800 hours and was completed by 1100 hours. There were two brief breaks during the battery testing period.

The battery was developed in consultation with Mr Geoffrey Marshall (Clinical Neuropsychologist, Royal North Shore Hospital) after extensive review of the previous literature. Several of the tests chosen for this study have been used in prior studies, although some are unique to this data. The exact order was determined to minimise any possible learning effect being carried over between tests.

The primary investigator was instructed in the administration of all tests and a standard instruction set was developed in order that all subjects be given the battery in a uniform manner. The final five tests were administered by Mr Marshall given their added complexity. The primary investigator acted as observer in the second sequence.

The absence of any pre-existing knowledge of the magnitude of sleep disordered breathing was felt to be crucial to the blinding of the investigators in an effort to avoid administration bias. Furthermore, the selection of several short tests of various complexity immediately following PSG was determined to be the construct most likely to avoid the pitfalls of night to night variability of unmeasured domiciliary sleep quality. Finally, the use of short tests was employed to minimise the subjective potential for inattention secondary to repetition.

2.6 Statistical Methods for Analysis of Neurocognitive Outcomes.

2.6.1 Simple Regression Modelling.

All neurocognitive outcomes were normed according to the techniques discussed in Chapters 2.1.1 - 2.1.8. The t-scores, z-scores or percentiles were then used for the analysis of the data by simple regression. The outcomes for arousal indices and hypoxic variables were considered in turn.

All cognitive tests were separately entered as dependent variables. The various parameters of sleep disturbance were entered as independent variables and all derived p values < 0.05 were considered statistically significant and are presented in the results.

2.6.2 Stepwise and Multiple Regression Techniques.

Having defined the relationships of statistical significance by simple linear regression, stepwise regression was then performed between each of the cognitive outcomes, about which there were a priori hypotheses, and all measures of sleep disordered breathing in order to ascertain which of the relationships were statistically independent.

From the stepwise regression analysis, several independent interrelationships between cognitive and sleep variables were able to be demonstrated. In order to ensure that no important explanatory variable was omitted from the final analysis, further multiple regression was performed. Each defined cognitive variable was entered as the dependent variable. Independent variables of sleep disordered breathing identified in the stepwise regression, were forced into the multiple regression model and omitted variables entered individually to

determine their effect over and above variables identified from the preceding stepwise regression. The SILS V_t was forced into each equation to control for the effect of general intellectual ability. All steps of the regression modelling are presented in the results.

2.6.3 Secondary Analysis of the Data.

Secondary analyses of clinically relevant outcomes, derived by stepwise and multiple regression, were performed by stratifying the normed cognitive data into quartiles of performance. Similarly, each relevant measure of sleep disordered breathing was stratified into quartiles. The cut-offs for determining percentile decrements in performance, for a given degree of sleep disturbance, were determined by calculation of the sensitivity and specificity curves at each interval. The actual value for each measure of SDB was derived by inspection of the quartile ranges. The odds ratios and confidence intervals for determining the risks of impairment at a given respiratory / arousal index were calculated from the logit transformed probability scale using the asymptotic standard error.

2.6.4 Validation Procedure for the Modified Arousal Scoring Method.

Having re-scored the arousals according to the modified criteria, inter and intra rater variability was determined by paired t-testing of the total number of arousals within each category. A statistically non-significant difference was considered evidence of satisfactory inter-rater agreement. A satisfactory level of concordance was considered to have been achieved for both intra and inter-rater scoring at an r-value of > 0.95 . This was derived by linear regression of the total number of each arousal type between the eight rescored studies of the primary investigator and the eight studies which were rescored by the second observer.

The blinding code was then broken, and the individual totals for each arousal type were assigned to the subjects in the study. All previous steps of analysis from single linear to stepwise and multiple regression were performed using the modified arousal index (AI-M) of interest as the dependent variable and all cognitive outcomes as independent variables. All outcomes of analysis are presented in the results.

2.7 Melatonin Samples.

2.7.1 Sample Collection Methods.

All samples were collected by the chief investigator. Following enrollment, all subjects were given two 4 litre urine sample bottles. One teaspoon of boracic acid was added to each sample bottle to reduce odour and to act as a preservative. Subjects were asked to empty their bladders at 7pm on the night of polysomnography. All urine produced overnight through to 0700 the following morning was collected. They were then furnished with a second sample bottle for daytime collection and instructed to return the bottle to the Sleep Unit at the completion of the study i.e. 7pm that evening. Each of the two consecutive 12 hour samples were weighed to measure volume and 10mls stored at minus 70°C for subsequent analysis. Calculation of serum creatinine was performed on the night of PSG and urinary creatinine was determined from each sample. Body mass index was calculated in the standard fashion from measurements performed on the study night.

2.7.2 Analysis Methods.

2.7.2.1 Urinary 6 Sulphatoxymelatonin Analysis (6-SM).

All samples were analysed by radioimmunoassay¹⁶⁷ in the laboratory of Dr David Kennaway (Department of Obstetrics and Gynaecology Adelaide University). The night time samples were diluted prior to the running of the assay. All samples were run in duplicate. Three samples were lost to analysis as a result of unreturned daytime collection.

Total 12 hour urinary 6-SM was calculated by multiplying the measured 6-SM (nmols/L) by the calculated urinary volume (L). To calculate the urine volume,

the Siersbaek-Nielsen¹⁶⁸ nomogram was used. The nomogram determines the urinary output based upon the urinary creatinine, serum creatinine and the body mass index. This step was deemed necessary to control for the possibility of missed samples during the course of collection.

2.7.2.2 Polysomnographic Analysis of Hypoxia and Arousal.

All sleep studies were scored according to R & K ⁸³ criteria. Measures of hypoxia including minimum saturation in REM and NREM, RDI in REM and NREM, the total RDI and the arousal index were extracted in the same manner as those described in Chapter 2.4.2 for the cognitive data set.

2.7.3 Statistical Analysis of Data.

The diurnal and nocturnal results were separately analysed. All values were correlated by simple linear regression against the variables of sleep disordered breathing. A p-value of < 0.05 was used to determine relationships of statistical significance.

A subanalysis was also performed to compare subjects without significant OSA (RDI < 5) and those with moderate OSA (RDI > 20). Unpaired t-testing of the subjects within each group was performed to examine whether an intergroup relationship existed external to the simple linear correlation. Results are presented in the subsequent chapter.