

DISCUSSION CHAPTER

4.1 The role of melatonin in the genesis of daytime sleepiness.

4.1.1 Outcome of the Study to examine total melatonin production.

The results of the experiment performed did not demonstrate any significant correlation between the production of melatonin (as measured by urinary metabolite) and any variable of sleep disordered breathing. Despite the hypothesis reasonably suggesting a possible physiological link, none was determined. In an effort to discriminate between normal subjects ie those with a total RDI < 5, and moderate OSA subjects (RDI > 20) unpaired t-testing on both daytime and nighttime samples failed to demonstrate any statistically significant difference. Inspection of the data spread and examination of the p and r values of the analyses would suggest that the absence of a positive finding was not related to insufficient numbers and that increasing sample size would be unlikely to alter the outcome.

In summary, there is no evidence from the data presented, to support the hypotheses that there is excess production of melatonin secondary to sleep disordered breathing. It is therefore unlikely that melatonin production as measured in this manner, contributes to daytime sleepiness in OSA.

The findings of this study are in keeping with those of Entzian et al¹³⁵ who similarly failed to demonstrate abnormal melatonin profiles in OSA patients versus normals.

4.1.2

Limitations of the study.

Although all steps to exclude confounding variables (age, alcohol consumption, medications and other serious medical conditions) were taken, formal drug screening and blood alcohol levels were not performed, and may have been useful in excluding some subjects.

The question of whether the overnight profile of melatonin secretion in OSA is abnormal was not examined. However, the previous data of Entzian et al would not support this hypothesis, at least not in the small numbers which were studied.

Although all steps to calculate urinary volumes by indirect methods were taken, it is possible that either the indirect measurements or the actual volumes were inaccurate which may have confounded the data. In fact simple linear regression between measured volumes and indirect calculated volumes demonstrate r-values > 0.85 for both diurnal and nocturnal measures suggesting reasonable patient compliance and making the possibility of error in total excretion seem unlikely.

Finally, the selection of patients from a group presenting to a Sleep Disorders Unit may suggest an element of selection bias. This may confound the data if indeed all subjects were being studied on the basis of excessive sleepiness. However, many of the subjects had little in the way of sleepiness but were being studied primarily for investigation of snoring and/or witnessed apnoeic events.

The alternate protocol of using matched normals was considered, but the limitations of the laboratory made this difficult and would have substantially extended the study time. While it is reasonable to accept that an alternate outcome may have been determined, the investigators were primarily interested in the patient group presenting to a Sleep Unit and in that setting the negative finding remains valid.

4.1.3 Further investigations.

It is evident from the previous chapter that several alternate studies could be undertaken to further explore the primary hypothesis. It would be ideal to measure either continuous or hourly serum levels of melatonin over the course of the night to demonstrate whether there is measurable phase shift. Additionally, the use of normal controls matched for age, sex and health status should be considered. It is likely that moderate numbers would be required and access to accurate serum assay techniques would be essential.

4.2 The relationship between variables of sleep disordered breathing and cognitive function.

4.2.1 The effects of sleep fragmentation on immediate and delayed short term memory.

The outcome of this study clearly demonstrates a relationship between sleep fragmentation, as measured by the standard arousal criteria, and decrements in both immediate and delayed short term memory. The relationship, by multiple regression, has been demonstrated to be independent of the total RDI and the NREM RDI when general intellectual ability is entered as a controlled variable. This finding is supported by the work of Telakivi⁷⁷ who confirmed that sleep fragmentation with mild hypoxia and daytime somnolence were the main causes of memory dysfunction. Unfortunately, while the group had been adjusted for age and obesity to exclude confounders, there is no evidence that they had been controlled for levels of education.

Other confirmatory evidence can be found in the previous work of Bedard et al⁸⁵. In a small study of only 20 patients, they demonstrated progressive deficits in

immediate and delayed recall in moderate to severe OSA patients. Unfortunately, no quantification of sleep fragmentation was reported and no data to define the magnitude of the insult required for a given deficit was demonstrated.

The relationship between sleep fragmentation and memory dysfunction is significant. The large sample size, the construct of using normal data and controlling for intellectual ability have allowed useful analysis with respect to clinical outcomes to be determined (see Chapter 4.2.4). The regression analyses which confirm this finding to be independent of other sleep disordered variables, in particular the total RDI is one of major clinical importance. For the group of patients with repetitive arousals in whom there is insufficient desaturation or apnoeas of less than 10 seconds, it demonstrates that they may well have cognitive dysfunction without sleep apnoea as defined by standard criteria.

4.2.2 The effect of REM related OSA on information processing speed.

The findings of a discrete relationship between sleep apnoea in REM and a decline in information processing speed is both novel and clinically pertinent.

To date there have been no reports to confirm this relationship, although most studies have not separated NREM and REM respiratory events for separate analysis. The relationship is independent of the arousal index, the amount of apnoea in NREM sleep and the total RDI.

It appears that disruption of REM sleep by repetitive apnoea gives rise to a unique decrement in information processing speed. The clinical importance of this finding is most obvious for the common group of OSA patients with predominantly REM OSA in whom the total RDI and the arousal indices may be only mildly elevated. The question of whether this patient group require treatment has until now been very

unclear. The findings reported in this study would suggest that these subjects are at risk of cognitive decline and this may colour clinical decision making.

4.2.3 The relationship between sleep hypoxaemia and vigilance.

This study has demonstrated a relationship between vigilance (as measured by Steer Clear) and the magnitude of the hypoxic injury as measured by the log of the time with a saturation < 90%. Previous work by Flemons et al⁹⁹ using the same simulator found no relationship between SC hits and the severity of desaturation as measured by the lowest recorded saturation. Indeed, this study also failed to demonstrate a correlation with this parameter in either REM or NREM sleep. This suggests it is not the minimum level which is critical in determining vigilance decrement, but the total time of the hypoxic insult.

Of note is the augmentation of the relationship when Vt (measure of general intellectual ability) is entered into the regression analysis. This supports the argument to control for this variable in studies which relate to cognitive outcomes and suggests that general intellectual ability may play a part in diminishing the effect of hypoxia on vigilance. This may help to explain the inability of previous data to demonstrate similar relationships.

The addition of the arousal index to the multiple regression equation further increases the statistical significance of the relationship. This augmentation would support the hypothesis that while hypoxia directly contributes to vigilance decrement, the addition of sleep fragmentation to hypoxia magnifies the impact of this insult on the ability of the individual to maintain vigilance.

4.2.3.1 Implications for driving.

The work of Flemons et al⁹⁹ and the previous studies by Findley et al^{97, 98} have attempted to demonstrate a relationship between motor vehicle accidents (MVA) and OSA. However, none of the studies were able to demonstrate any single respiratory variable as being primarily related to accident rates. The more recent work by George et al^{100, 101} demonstrated more tracking errors in OSA patients versus normals. This error in tracking is only weakly associated with sleepiness as measured by MSLT and overall the AHI and the MSLT explain less than 25% of the variance. The findings of the previous Chapter 4.2.3, would appear to be unique in identifying a strong correlation between a single sleep variable and the number of collisions on Steer Clear. Furthermore, the addition of sleep fragmentation to the model increases the strength of the relationship. This study has not attempted to define thresholds for hypoxia or arousal which may delimit the risk for a given number of S. C. collision errors as this may not accurately reflect true driving risk. There is no data which directly correlates Steer Clear errors to the expectant risk of an MVA and as such it may not be the ideal tool for defining MVA risk in OSA. The findings do however suggest that specific variables of sleep disordered breathing may be able to be individually identified which in turn may be used to define the risk. This work, for the first time identifies one possible factor, although the suggestion from the work of Flemons & George is that it is likely to be multifactorial.

4.2.3.2 Further investigations.

Given the findings outlined above, it would be useful to develop a tool which incorporates both the reaction time component of Steer Clear with the tracking error function of the Divided Attention Driving Test. The need for large numbers of matched normals and the imperative for controlling for general intellectual ability are evident. The relationship between errors of judgement, reaction time and parameters

of sleep disordered breathing are likely to be important in defining driving risks but will need to be measured in parallel with actual on road accidents to be meaningful.

4.2.4 Implications for intention to treat. Interpretation of secondary analysis.

The most pertinent clinical findings of the study were determined to be the relationship between the arousal index and short term memory and REM sleep apnoea and information processing. This does not diminish the value of the other findings such as the relationship between Type III & IV arousals and working memory (see Chapter 4.3.2) or that of hypoxia and vigilance. However, since both the AI and the REM RDI are data derived by most clinical sleep laboratories, it was felt that they may have the most immediate value in the clinical setting.

Following the methods of analysis described in Chapter 2.6.3, and given the strength of correlation between AI and Logical Memory 1 & 2, the data was re-analysed to derive the relative risk of performing at or below a given quartile for a given degree of sleep disordered breathing. Thus it can be said that at an arousal index of 37 per hour there is a 2.35 chance of an individual performing a test of immediate short term memory in the lowest quartile. At an arousal index of 34 per hour, they are 5 times more likely to perform a test of delayed short term memory in the lowest quartile. It is important to recognise that all data was normed according to age and level of education prior to analysis which would exclude the confounding variable of skewed distribution. Indeed, by dividing the values for each test into quartiles, all values are re-distributed in a Gaussian fashion further diminishing the effect of skewed performance as a confounder. Of note, the mean arousal index for the group is 29/hour suggesting that a moderate proportion of all patients attending a sleep clinic are likely to demonstrate a modicum of short term memory decrement. In as much as the 55 subjects were all randomly selected, it is likely that they are representative of the group as a whole.

At a REM RDI of 37 per hour a subject is 2.7 times more likely to perform a test of information processing speed in the lowest quartile. Furthermore, at a REM RDI of 14/hour, they are 3.2 times more likely to perform at or below the 50% percentile. To ensure that this outcome was not a product of the quantity of REM sleep present, further analyses were performed which demonstrated that the relationship between the total number of respiratory events in REM and information processing speed remains significant and independent of the time spent in REM sleep and the other variables of sleep disordered breathing (AI, total RDI, and the log of the desat < 90%).

Thus while a REM RDI of 15 per hour, with a normal total RDI would be considered by many to be trivial OSA, this data demonstrates this may not be so. In fact, it may indicate a new threshold for considering therapy particularly if the subject is symptomatic.

4.2.5 Study Limitations.

A summary of the major studies to-date demonstrates several flaws in design. Clearly the timing of the test battery is important both to ensure blinding by the administrators but also to ensure that direct correlations to the degree of sleep disturbance can be made. Difficulties in matching OSA patients for level of education as well as age, sex and pre-morbid cognitive function need to be considered. Sufficient numbers are required if a clinically pertinent finding is to be extrapolated to the OSA population as a whole, to define an 'intention to treat' decrement in cognitive function. Furthermore, given the diversity of tests used in the various batteries described above, there would not appear to be any single tool which could be universally applied to assess cognitive abnormality in OSA.

Given the limitations of previous studies in the area, this study has been developed in an effort to more rigorously blind the test administrators. The tests were chosen and administered in such a manner as to accomplish the greatest yield in a relatively short administration time to avoid deteriorating function due to fatigue. Furthermore, all tests were delivered immediately following PSG to ensure that the results were able to be directly linked to the magnitude of sleep disturbance on the prior night. All cognitive variables were normed against published values for each test and the effect of general intellect was controlled for in the analysis to ensure all subjects could be considered on par.

Despite these efforts, several possible limitations within the study are evident. Given the night-to-night variability of OSA¹⁶⁹, one cannot be certain that all subjects were studied on a night which was representative of the usual magnitude of their sleep apnoea. Confounders such as increased alcohol at home cannot be accounted for in this analysis. The study may well have been further strengthened by confirmatory domiciliary PSG on one or possibly more nights. This facility however, is resource intensive and complex. While it may have provided additional information it is unlikely that it would diminish the findings as presented.

The concomitant issues of microvascular disease or alcohol related brain damage cannot be entirely dismissed. Although every effort was made to screen patients for confounding intercurrent illnesses, blood alcohol screening and CT scans of the brain were not performed primarily for reasons of time and cost.

The single most important criticism of this work lies in both the number and low *r*-squared values for the correlations. There is no doubt that 17 correlations out of 147 possible analyses is small indeed. One would have imagined that potentially more positive findings may have surfaced given the previous literature suggesting relationships between OSA and cognitive dysfunction. However, it was hoped that the use of stepwise regression to ascertain by filtering which of the relationships was

most significant, would diminish the impact of a Type I error. There are several possible explanations for both the poor number of positive outcomes and the low variance derived for each correlation.

1. The fashion in which sleep studies are scored remains arbitrary. As yet there is no universally accepted rule for defining respiratory events and little understanding regarding the inter-relationship between arousals and upper airway events. Indeed, even the degree of desaturation required for a respiratory event remains non-standard. Therefore, the sensitivity of the polysomnograph remains ill-defined as a tool for determining the total insult of OSA in this setting. Perhaps alternate methods of scoring events, both respiratory and cortical are needed to improve the correlation between OSA and cognitive decrement.

2. The tools used for this experiment have been largely developed for use in disease states outside of OSA, including stroke and closed head injury. They have relatively large standard errors about the mean, and do not necessarily define the exact derivation of the normal group which may be critical in appraising the sensitivity of these tools. At best there is a poor correlation between OSA and cognitive decrement, as determined by the r-squared values in this work, suggesting that a direct cause and effect relationship between OSA and cognitive dysfunction is ill-defined. There are likely to be several co-factors including motivation, general well being and sleepiness which are at best poorly measured both here and in general.

3. Several of the tests used the experimental paradigm clearly overlap in terms of the outcome being measured. Indeed, simple inattention to the test instructions can impair the performance of the task assigned. While this work has sought to define the most important measurement parameter, it is fair to say that this is arbitrary and perhaps a little unrealistic. However, there are no current validated tools which can simply measure a single cognitive function in isolation.

Development of such tools may improve the ability to detect decrement in OSA patients.

4. The selection of tests was determined in consultation with a clinical neuropsychologist and based upon a review of all of the previous literature available in the area. The majority of the tests were developed in Europe or America as were the derived normative tables. It is an accepted criticism that Australian populations may differ in performance accounting for some of the lack of positive correlations found in this experiment. This unfortunately is a weakness which could not be overcome.

It is possible that further enrollment may have established other relationships beyond those demonstrated. However, since interim analysis demonstrated adequate power, the study was terminated. Analysis of the non-significant results demonstrates very low r-values and suggests that larger numbers would be very unlikely to have defined further relevant relationships. Having achieved a pre-determined power of $\geq 90\%$ at 55 subjects, Type II error is deemed unlikely to have confounded the study outcomes.

In summary, the possibility of deriving further relationships based upon test selection is an issue for consideration. A substantial range of tests have been used in this and previous studies looking for ways to define the cognitive insult in OSA. The tests used in this battery were selected for simplicity and brevity. It is possible that more sensitive and sophisticated tools may have demonstrated greater outcomes although the tools used for this work are the standard instruments of clinical neuropsychology. It is hoped that these somewhat blunt instruments, having demonstrated the abnormalities described, will be used to develop better tools for assessing cognitive dysfunction in OSA, which is a predictably more subtle disease than dementia or closed head injury.

4.3 The effects of sleep fragmentation, as defined by modified criteria, on outcomes of cognitive function.

4.3.1 The effects of brief (Type I) and intermediate (Type II) duration arousals on immediate and delayed short term memory (STM).

Despite the evidence of a strong correlation between the AI scored by standard criteria (AI-S) and outcomes of immediate and delayed STM, the division of arousals into brief and intermediate failed to demonstrate such a relationship. Indeed, the Type I arousals did not correlate with any outcomes of short term memory, although a relationship to subjective sleepiness (ESS) was demonstrated. This outcome was also seen in the simple linear regression of ESS against AI-S. The r-value of 0.27 would imply that brief arousals make at least a minor contribution to the genesis of daytime somnolence. This finding is supported by the work of Martin et al²¹ who demonstrated a similar increase in somnolence on MSLT in an experimental sleep fragmentation model. Indeed, the arousals in Martin's study were at least three seconds in duration and would suitably mimic the Type I and/or Type II arousals of this study.

The lack of any definable relationship between Type I arousals and outcomes of short term memory suggests that it is not necessarily the brief arousals of sleep which determine the short term memory decrement which has been described. Furthermore, while Type II arousals do correlate with a decline in immediate short term memory, there is a non-significant relationship with delayed short term memory.

Simple linear regression of the AI-S versus Type I and Type II arousals demonstrate very significant correlations (Type I $p < 0.0001$, $r^2 = 0.68$, Type II $p < 0.0001$, $r^2 = 0.30$). Despite this, as separate entities neither arousal sub-type is able to generate the same degree of significance when correlated with either immediate or delayed

short term memory. The implications of these findings are that there is likely to be some other component of sleep fragmentation, when measured by arousals scored by standard R & K criteria, which determines the short term decrement previously demonstrated in this work. At this point, there is no obvious evidence within the data analysis to suggest a likely contributor.

4.3.2 Effects of shifts in sleep stage (Type III) and arousals to wakefulness (Type IV) on working memory.

The findings of a very significant relationship between these two arousal sub-types and decrements in working memory is both novel and potentially clinically relevant.

The data supports the hypothesis that repetitive shifts to a lesser stage of sleep and repetitive wakefulness correlates with a decline in working memory and that this is independent of all major variables of sleep disordered breathing (AI-S, log desat < 90%, Total RDI). Of note, unlike the Type I and II arousals, these arousals demonstrate a much lesser internal correlation with the AI-S (Type III $p = 0.89$, Type IV $p < 0.05$, $r^2 = 0.17$).

This new data may have implications for clinical entities other than OSA, particularly insomnia and in the elderly whose usual sleep is characterised by poor efficiency, inability to sustain deep sleep and repetitive awakenings. This data may also suggest plausible reasons for poor performance amongst shift workers and on-call medical staff who demonstrate similar sleep fragmentation and cognitive decline^{58, 170}.

The original paper by Stepanski et al³⁸ did not, unfortunately, incorporate cognitive outcomes although it did confirm that arousals to wakefulness did not correlate with daytime sleepiness, a finding in keeping with this data.

The only corroborative data which supports the finding of this study is the earlier work of Cheshire et al⁸⁰. They demonstrated a similar relationship between the mean shifts in sleep stage/hour and a decrement in PASAT performance. Unfortunately the study used a subset of more moderate OSA patients (AHI > 15/hour plus symptoms) in whom there are fewer sleep stage shifts and wakeful

periods as demonstrated by the work of Stepanski. It therefore supports the proposed hypothesis but may have underestimated the true strength of the relationship due to smaller numbers and selection bias.

4.3.3 Implications for sleep disordered breathing.

The data presented in this work with respect to arousal subscore may have limited application for sleep disordered breathing. The magnitude of the relationship between standard arousal scoring and cognitive decrement is clearly of greater importance than the results of this sub-classification. As Stepanski³⁸ demonstrated, sleep apnoeic patients demonstrated fewer Type III and IV arousals although both this study and that of Cheshire et al confirms their relevance to working memory decrement. The findings of this analysis may, however, have a bearing on cognitive performance within other sub-groups of sleep disorders. Patients with insomnia, in whom there is an absence of demonstrable sleep disordered breathing, have been shown to perform less well than normals^{78, 171, 172}. It is conceivable that cognitive decrement is a function of poor sleep efficiency and repetitive wakefulness in this group, a hypothesis which deserves further investigation.

4.3.4 Limitations of the modified arousal scoring procedure.

Although the criteria for the modified arousal scoring were based on the previous published works of Stepanski and Cheshire, substantial corroborative evidence is missing. The work presented here demonstrates similar but not identical findings to the two previous studies. The variations in outcomes probably reflects the differing sample sizes and patient selection between this and the previous work.

The lack of correlation between the lesser indices of arousal and short term memory decrement or indeed Type III or IV arousals and this outcome may be seen as

inconsistent with the work of Martin et al^{21, 173}. However, the tests of memory decrement (logical memory I and II) were not utilised by these authors and the arousals of these previous studies were experimentally induced to mimick OSA sleep fragmentation rather than being related to possible upper airway events. Thus a true comparative analysis of this work and the previous data is not possible.

Finally, the data have not been subjected to subclassification between NREM and REM sleep as have much of the other analyses in this work. At this point it remains a plan for further investigation and ongoing research.

4.4 Final summary of the experimental data.

It has been a long and often arduous journey to have arrived at the completion of this study.

Although many of the outcomes were predicted on the basis of previous work, many appear to be both new and deserving of further investigation.

It is the belief of this author that much about the genesis of cognitive dysfunction in sleep disordered breathing can be explained by the analyses performed on this data set. The enrollment of adequate numbers and a study construct which allowed the development of new treatment thresholds may have much worth for all the workers in this area. Unexpected findings such as that between major arousals and working memory as well as that between vigilance and the total hypoxic insult will hopefully raise new questions regarding functional daytime impairment and sleep disturbance. This author hopes to continue to seek answers to these questions and hopefully develop more sensitive tools and techniques for characterising the numerous cognitive decrements which have their genesis in sleep.